# PHYSIOLOGY OF THE GASTROINTESTINAL TRACT

FOURTH EDITION

Volume 1

Editor-in-Chief • Leonard R. Johnson

Associate Editors

Kim E. Barrett • Fayez K. Ghishan
Juanita L. Merchant • Hamid M. Said
Jackie D. Wood

*PHYSIOLOGY OF THE GASTROINTESTINAL TRACT*

**FOURTH EDITION**

**This Page Intentionally Left Blank**

# Physiology

## of the

# Gastrointestinal Tract

**FOURTH EDITION**

*Volume 1*


Editor-in-Chief

**Leonard R. Johnson**
*Department of Physiology*
*University of Tennessee College of Medicine*
*Memphis, Tennessee*

Associate Editors

**Kim E. Barrett**
*Department of Medicine*
*University of California, San Diego, School of Medicine*
*University of California San Diego Medical Center*
*San Diego, California*

**Juanita L. Merchant**
*Departments of Internal Medicine and Molecular and Integrative Physiology*
*University of Michigan*
*Ann Arbor, Michigan*

**Fayez K. Ghishan**
*Department of Pediatrics*
*Steele Children's Research Center*
*University of Arizona Health Sciences Center*
*Tucson, Arizona*

**Hamid M. Said**
*Departments of Medicine and Physiology/Biophysics*
*University of California*
*Irvine, California,*
*and VA Medical Center-151*
*Long Beach, California*

**Jackie D. Wood**
*Departments of Physiology and Biology and Internal Medicine*
*The Ohio State University College of Medicine*
*Columbus, Ohio*



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier


Elsevier Academic Press
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
84 Theobald's Road, London WC1X 8RR, UK

This book is printed on acid-free paper. ♾



**Library of Congress Cataloging-in-Publication Data**

Physiology of the gastrointestinal tract / editor-in-chief, Leonard R. Johnson ; associate editors, Kim Barrett ... [et al.].-- 4th ed.
p. ; cm.
Includes bibliographical references and index.
ISBN 0-12-088394-5 (set : alk. paper) -- ISBN 0-12-088395-3 (v.1 : alk. paper)
ISBN 0-12-088396-1 (v.2 : alk. paper)
1. Gastrointestinal system--Physiology. I. Johnson, Leonard R., 1942-
[DNLM: 1. Gastrointestinal Tract--physiology. WI 102 P578 2006]
QP145.P492 2006
612.3′2--dc22 2005043533

**British Library Cataloguing in Publication Data**
A catalogue record for this book is available from the British Library

Volume 1: ISBN 13: 978-0-120883950 • ISBN 10: 0-12-088395-3
Volume 2: ISBN 13: 978-0-120883967 • ISBN 10: 0-12-088396-1
Two Volume Set: ISBN 13: 978-0-120883943 • ISBN 10: 0-12-088394-5

For all information on all Elsevier Academic Press publications
visit our Web site at www.books.elsevier.com

Printed in the United States of America

06 07 08 09 10 9 8 7 6 5 4 3 2 1

# Contents

## Section II. Neural Gastroenterology and Motility

## Section V. Digestion and Absorption

**This Page Intentionally Left Blank**

# Contributors

**Nada Abumrad,** Departments of Medicine and Nutritional Sciences, Washington University, Campus Box 8031, 660 South Euclid Avenue, Saint Louis, Missouri 63110

**Yasutada Akiba,** Brentwood Biomedical Research Institute, Building 114, Suite 217, West Los Angeles VAMC, Los Angeles, California 90073

**James M. Anderson,** Department of Cell and Molecular Physiology, University of North Carolina at Chapel Hill, 6312 MBRB, 103 Mason Farm Road, Chapel Hill, North Carolina 27599-7545

**Nancy C. Andrews,** Department of Basic Sciences and Graduate Studies, Harvard Medical School, and Department of Pediatrics, Children's Hospital, Howard Hughes Medical Institute, Gordon Hall, 25 Shattuck Street, Boston, Massachusetts 02115

**Barry E. Argent,** Institute for Cell and Molecular Biosciences, University Medical School, Newcastle upon Tyne NE2 4HH, United Kingdom

**Qasim Aziz,** University of Manchester, Clinical Sciences Building, Hope Hospital, Stott Lane, Salford M6 8HD, United Kingdom

**Liqun Bai,** Department of Pediatrics, Steele Children's Research Center, University of Arizona Health Sciences Center, 1501 North Campbell Avenue, Tucson, Arizona 85724-5073

**Longchuan Bai,** Department of Internal Medicine, University of Michigan, 1150 West Medical Center Drive, 3510 MSRB I, Ann Arbor, Michigan 48109

**Graham S. Baldwin,** Department of Surgery, University of Melbourne, Austin Health, Heidelberg, Victoria 3084, Australia

**Kim E. Barrett,** Department of Medicine, University of California, San Diego, School of Medicine, University of California San Diego Medical Center 8414, 200 West Arbor Drive, San Diego, California 92103-8414

**John A. Barnard,** Department of Pediatrics, Division of Molecular Medicine, The Ohio State University College of Medicine, Columbus Children's Hospital, 700 Children's Drive WA2011, Columbus, Ohio 43205

**Marc D. Basson,** John D. Dingell Veterans Affairs Medical Center, Department of Surgical Service, Wayne State University, Detroit, Michigan 48201

**Michael J. Beyak,** Gastrointestinal Diseases Research Unit, Queen's University, Kingston, Ontario, Canada, and Department of Biomedical Science, University of Sheffield, Sheffield S10 2TN, United Kingdom

**Richard S. Blumberg,** Department of Medicine, Harvard Medical School, and Division of Gastroenterology, Hepatology, and Endoscopy, Harvard Digestive Diseases Center, Laboratory of Mucosal Immunology, Thorn 1419, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115

**Guido T. Bommer,** Division of Molecular Medicine and Genetics, Department of Internal Medicine, University of Michigan School of Medicine, LSI 5-183A, 210 Washtenaw Avenue, Ann Arbor, Michigan 48109-2216

**Simon J. H. Brookes,** Department of Human Physiology and Centre for Neuroscience, Flinders University, Bedford Park, South Australia 5042

**Richard S. Bruno,** Department of Nutritional Sciences, University of Connecticut, 3624 Horsebarn Road Ext, Unit 4017, Storrs, Connecticut 06269-4017

**David C. E. Bulmer,** Department of Biomedical Science, University of Sheffield, Sheffield S10 2TN, United Kingdom

**Nigel W. Bunnett,** Departments of Surgery and Physiology, University of California, San Francisco, Room C317, 521 Parnassus Avenue, San Francisco, California 94143-0660

**Jenny M. Buzan,** Digestive Health Center of Excellence, University of Virginia, Charlottesville, Virginia 22908-0708

**R. Maynard Case,** Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, United Kingdom

**Zhouji Chen,** Division of Endocrinology, Metabolism, and Lipid Research, Washington University School of Medicine, Campus Box 8046, 660 South Euclid Avenue, Saint Louis, Missouri 63110

**Fedias Leontiou Christofi,** Departments of Anesthesiology and Physiology and Cell Biology, The Ohio State University, 226 Tzagounis Medical Research Facility, Columbus, Ohio 43210

**Steve Claypool,** Laboratory of Mucosal Immunology, Thorn 1419, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115

**Ray E. Clouse,** Departments of Medicine and Psychiatry, Washington University, 660 South Euclid Avenue, Campus Box 8124, Saint Louis, Missouri 63110

**Steven M. Cohn,** Digestive Health Center of Excellence, University of Virginia, 2 Jefferson Park Ave, Room 2091, Charlottesville, Virginia 22908-0708

**James F. Collins,** Department of Pediatrics, Steele Children's Research Center, University of Arizona Health Sciences Center, 1501 North Campbell Avenue, Tucson, Arizona 85724-5073

**Helen Joan Cooke,** Department of Neuroscience, The Ohio State University, 4066D Graves Hall, 333 West Tenth Avenue, Columbus, Ohio 43210

**Marcello Costa,** Department of Human Physiology and Centre for Neuroscience, Flinders University, Bedford Park, South Australia 5042

**Robert J. Cousins,** Boston Family Professor of Nutrition, Center for Nutritional Sciences, University of Florida, PO Box 110370, Gainesville, Florida 32611-0370

**John Cuppoletti,** Department of Molecular and Cellular Physiology, University of Cincinnati College of Medicine, 231 Albert Sabin Way ML 0576, Cincinnati, Ohio 45267-0576

**Nicholas O. Davidson,** Division of Gastroenterology, Washington University School of Medicine, Campus Box 8124, 660 South Euclid Avenue, Saint Louis, Missouri 63110

**Paul A. Dawson,** Department of Internal Medicine, Section of Gastroenterology, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, North Carolina 27157

**Sarah A. De La Rue,** Digestive Health Center of Excellence, University of Virginia, Charlottesville, Virginia 22908-0708

**Nicholas E. Diamant,** Departments of Medicine and Physiology, University of Toronto, 6B Fell 6-176, Toronto Western Hospital, 399 Bathurst Street, Toronto, Ontario, Canada M5T 2S8

**Bonny Dickinson,** Department of Pediatrics, Research Institute for Children, Children's Hospital, Research and Education Building Room 2231, 200 Henry Clay Avenue, New Orleans, Louisiana 70118

**Chris J. Dickinson,** Division of Pediatric Gastroenterology, University of Michigan, D3252 MPB, 1500 East Medical Center Drive, Ann Arbor, Michigan 48109-0718

**Graham J. Dockray,** Department of Physiology, School of Biomedical Sciences, Crown Street, University of Liverpool, PO Box 147, Liverpool, L69 3BX United Kingdom

**Daniel J. Drucker,** Banting and Best Diabetes Centre, Toronto General Hospital, University of Toronto, 200 Elizabeth Street MBRW 4R-902, Toronto, Ontario, Canada M5G 2C4

**Pradeep K. Dudeja,** Department of Medicine, University of Illinois at Chicago College of Medicine, Research and Development, Jesse Brown VAMC, MP151, 820 South Damen Avenue, Chicago, Illinois 60612

**Alexandrine During,** Department of International Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205

**Lars Eckmann,** Department of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0665

**Paul Enck,** Department of Internal Medicine VI, University Hospitals Tuebingen, Osianderstr. 5, 72076 Tuebingen, Germany

**Ella W. Englander,** Department of Surgery, The University of Texas Medical Branch, 815 Market Street, Galveston, Texas 77550

**B. Mark Evers,** Department of Surgery, The University of Texas Medical Branch, 301 University Boulevard, Galveston, Texas 77555-0536

**Eric R. Fearon,** Division of Molecular Medicine and Genetics, Departments of Internal Medicine, Human Genetics, and Pathology, University of Michigan School of Medicine, LSI 5-183A, 210 Washtenaw Avenue, Ann Arbor, Michigan 48109-2216

**John G. Forte,** Department of Molecular and Cell Biology, University of California, 241 LSA, MC 3200, Berkeley, California 94720

**Mark R. Frey,** Department of Pediatric Gastroenterology, Hepatology, and Nutrition, S4322 MCN Vanderbilt University, Nashville, Tennessee 37232-2576

**Vadivel Ganapathy,** Department of Biochemistry and Molecular Biology, Medical College of Georgia, 1459 Laney-Walker Boulevard, Augusta, Georgia 30912-2100

**Cheryl E. Gariepy,** Division of Pediatric Gastroenterology, University of Michigan, D3252 MPB, 1500 East Medical Center Drive, Ann Arbor, Michigan 48109-0718

**Michael D. Gershon,** Department of Anatomy and Cell Biology, Columbia University, College of Physicians and Surgeons, 630 West 168 Street, New York, New York 10032

**Fayez K. Ghishan,** Department of Pediatrics, Steele Children's Research Center, University of Arizona Health Sciences Center, 1501 North Campbell Avenue, Tucson, Arizona 85724-5073

**Andrew S. Giraud,** Department of Medicine, University of Melbourne, Western Hospital, Footscray 3011, Australia

**Fred S. Gorelick,** Department of Medicine, VA Healthcare CT, and Yale University School of Medicine, 950 Campbell Avenue, West Haven, Connecticut 06516

**D. Neil Granger,** Department of Molecular and Cellular Physiology, Louisiana State University Health Sciences Center, 1501 Kings Highway, Shreveport, Louisiana 71130-3932

**Michael A. Gray,** Institute for Cell and Molecular Biosciences, University Medical School, Newcastle upon Tyne NE2 4HH, United Kingdom

**George H. Greeley Jr.,** Department of Surgery, The University of Texas Medical Branch, 301 University Boulevard, Galveston, Texas 77555-0725

**Matthew B. Grisham,** Department of Molecular and Cellular Physiology, Louisiana State University Health Sciences Center, 1501 Kings Highway, Shreveport, Louisiana 71130-3932

**David Grundy,** Department of Biomedical Science, University of Sheffield, Sheffield S10 2TN, United Kingdom

**Michèle Gué,** Department of Physiology, Université Paul Sabatier, IFR31, Institut Louis Bugnard, BP 84225, INSERM U388, Laboratoire de Pharmacologie Moléculaire et Physiopathologie Rénale, 31432 Toulouse, Cedex 4, France

**Naren Gupta,** Department of Surgery, University of Virginia Health System, Charlottesville, Virginia 22908

**Earl H. Harrison,** Phytonutrients Laboratory, United States Department of Agriculture Human Nutrition Research Center, Building 307 C, Room 118 BARC-East, Beltsville, Maryland 20705

**William L. Hasler,** Department of Internal Medicine, Division of Gastroenterology, University of Michigan Health System, 3912, Taubman Center, Box 0362, Ann Arbor, Michigan 48109

**Gail Hecht,** Section of Digestive Diseases and Nutrition, University of Illinois, 840 South Wood Street, Room 738A (m/c 716), Chicago, Illinois 60612

**Mark R. Hellmich,** Department of Surgery, The University of Texas Medical Branch, 301 University Boulevard, Galveston, Texas 77555-0536

**Gerlinda E. Hermann,** Department of Neuroscience, Pennington Biomedical Research Center, 6400 Perkins Road, Baton Rouge, Louisiana 70808

**Bruce A. Hirayama,** Department of Physiology, The David Geffen School of Medicine at University of California Los Angeles, Los Angeles, California 90095-1751

**Anthony Hobson,** University of Manchester, Clinical Sciences Building, Hope Hospital, Stott Lane, Salford M6 8HD, United Kingdom

**Kim Hodges,** Section of Digestive Diseases and Nutrition, University of Illinois, 840 South Wood Street, Room 738A (m/c 716), Chicago, Illinois 60612

**Alan F. Hofmann,** Department of Medicine MC 0813, Division of Gastroenterology, University of California, San Diego, La Jolla, California 92093-0813

**Peter Holzer,** Department of Experimental and Clinical Pharmacology, Medical University of Graz, Universitätsplatz 4, A-8010 Graz, Austria

**V. Stephen Hunt,** Department of Molecular and Cellular Physiology, Beckman Center for Molecular and Genetic Medicine, Stanford University School of Medicine, Stanford, California 94305-5435

**Dawn A. Israel,** Division of Gastroenterology, Department of Medicine, Vanderbilt University Medical Center, Room 1012A MRB IV, 2215 Garland Avenue, Nashville, Tennessee 37232

**Claire Jacob,** Departments of Surgery and Physiology, University of California, San Francisco, Room C317, 521 Parnassus Avenue, San Francisco, California 94143-0660

**James D. Jamieson,** Department of Cell Biology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510

**Wen Jiang,** Department of Biomedical Science, University of Sheffield, Sheffield S10 2TN, United Kingdom

**Finn-Eirik Johansen,** Institute of Pathology, University of Oslo, Department of Pathology, Rikshospitalet University Hospital, Sognsvannsveien 20, N-0027 Oslo, Norway

**Leonard R. Johnson,** Department of Physiology, University of Tennessee College of Medicine, 894 Union Avenue, Memphis, Tennessee 38163

**Serhan Karvar,** Department of Molecular and Cell Biology, University of California, 245 LSA, MC 3200, Berkeley, California 94720

**Jonathan D. Kaunitz,** Department of Medicine, Division of Digestive Diseases, West Los Angeles VAMC, and UCLA School of Medicine, Building 114, Suite 217, Los Angeles, California 90073

**C. Keating,** Department of Biomedical Science, University of Sheffield, Sheffield S10 2TN, United Kingdom

**Stephen J. Keely,** Department of Medicine, University of California, San Diego, School of Medicine, University of California San Diego Medical Center 8414, 200 West Arbor Drive, San Diego, California 92103-8414

**Christopher G. Kevil,** Department of Pathology, Louisiana State University Health Sciences Center, 1501 Kings Highway, Shreveport, Louisiana 71130-3932

**Pawel R. Kiela,** Department of Pediatrics, Steele Children's Research Center, University of Arizona Health Sciences Center, 1501 North Campbell Avenue, Tucson, Arizona 85724-5073

**Sang Don Koh,** Department of Physiology and Cell Biology, University of Nevada School of Medicine, Reno, Nevada 89557

**Kris V. Kowdley,** University of Washington, Box 356174, 1959 NE Pacific Street, Seattle, Washington 98195

**Hanna Krysinska,** Child Health Research Institute, North Adelaide, 72 King William Road, South Australia 5006, Australia

**Rahul Kuver,** Department of Medicine, Division of Gastroenterology, University of Washington School of Medicine, Box 356424, 1959 NE Pacific Street, Seattle, Washington 98195

**Nicholas F. LaRusso,** Departments of Medicine, Biochemistry, and Molecular Biology, Mayo Clinic College of Medicine, 200 First Street Southwest, 1701 Guggenheim Building, Rochester, Minnesota 55905

**Sum P. Lee,** Department of Medicine, Division of Gastroenterology, University of Washington School of Medicine, Box 356424, 1959 NE Pacific Street, Seattle, Washington 98195

**Wayne I. Lencer,** Department of Pediatrics, Harvard Medical School, and Division of Pediatric Gastroenterology and Nutrition, Harvard Digestive Diseases Center, Gastrointestinal Cell Biology Laboratories, Enders 720, Children's Hospital Boston, 300 Longwood Avenue, Boston, Massachusetts 02115

**Rodger A. Liddle,** Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710

**Donald D. F. Loo,** Department of Physiology, The David Geffen School of Medicine at University of California Los Angeles, Los Angeles, California 90095-1751

**John Lynch,** Division of Gastroenterology, University of Pennsylvania School of Medicine, 415 Curie Boulevard, 600 Clinical Research Building, Philadelphia, Pennsylvania 19104

**Thomas Y. Ma,** Division of Gastroenterology and Hepatology, Departments of Medicine, Cell Biology and Physiology, Inflammatory Bowel Disease Program, University of New Mexico, MSC10 5550, 1 University of New Mexico, Albuquerque, New Mexico 87131-0001

**Wallace K. MacNaughton,** Department of Physiology and Biophysics, University of Calgary, 3330 Hospital Drive NW, Calgary, Alberta T2N 4N1, Canada

**Adhip P. N. Majumdar,** John D. Dingell Veterans Affairs Medical Center, Department of Medicine, Karmanos Cancer Center, Wayne State University, Research Service, Room-B-4238, 4646 John R, Detroit, Michigan 48201

**Gabriel M. Makhlouf,** Department of Medicine, Virginia Commonwealth University Medical Center, Sanger Hall, Room 12-003, 1101 East Marshall Street, Richmond, Virginia 23298

**Danuta H. Malinowska,** Department of Molecular and Cellular Physiology, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267-0576

**Robert G. Martindale,** Department of Surgery, Oregon Health and Science University, Portland, Oregon 97239

**Anatoliy I. Masyuk,** Department of Medicine, Mayo Clinic College of Medicine, 200 First Street Southwest, 1701 Guggenheim Building, Rochester, Minnesota 55905

**Tatyana V. Masyuk,** Department of Medicine, Mayo Clinic College of Medicine, 200 First Street Southwest, 1701 Guggenheim Building, Rochester, Minnesota 55905

**Gary M. Mawe,** Department of Anatomy and Neurobiology, The University of Vermont, D403A Given Building, 89 Beaumont Avenue, Burlington, Vermont 05405

**Kirk M. McHugh,** Department of Pediatrics, Division of Molecular Medicine, The Ohio State University College of Medicine, Columbus Children's Hospital, 700 Children's Drive WA2011, Columbus, Ohio 43205

**Juanita L. Merchant,** Departments of Internal Medicine and Molecular and Integrative Physiology, University of Michigan, 1150 West Medical Center Drive, 3510 MSRB I, Ann Arbor, Michigan 48109

**Donald J. Messner,** Bastyr University, 14500 Juanita Drive, NE, Kenmore, Washington 98028

**Steven M. Miller,** Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine, 200 First Street, SW, Rochester, Minnesota 55905

**Mulugeta Million,** CURE/Digestive Diseases Research Center and Center for Neurovisceral Sciences and Women's Health, Department of Medicine, Division of Digestive Diseases, David Geffen School of Medicine at the University of California, Los Angeles, and VA Greater Los Angeles Healthcare System, CURE Building 115, Room 117, 11301 Wilshire Boulevard, Los Angeles, California 90073

**Marshall H. Montrose,** Department of Molecular and Cellular Physiology, University of Cincinnati, Medical Science Building, Room 4253, 231 Albert Sabin Way, Cincinnati, Ohio 45267

**Timothy H. Moran,** Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Ross 618, 720 Rutland Avenue, Baltimore, Maryland 21205

**Karen F. Murray,** Children's Hospital and Regional Medical Center, University of Washington School of Medicine, 4800 Sand Point Way, NE, PO Box 5371/A5950, Seattle, Washington 98105

**Karnam S. Murthy,** Departments of Medicine and Physiology, Virginia Commonwealth University Medical Center, Sanger Hall, Room 12-003, 1101 East Marshall Street, Richmond, Virginia 23298

**W. James Nelson,** Department of Molecular and Cellular Physiology, Beckman Center for Molecular and Genetic Medicine, Stanford University School of Medicine, Stanford, California 94305-5435

**Philip T. Nowicki,** Departments of Pediatrics and Physiology, Center for Cell and Vascular Biology, Columbus Children's Research Institute, College of Medicine and Public Health, The Ohio State University, Children's Hospital, 700 Children's Drive, Columbus, Ohio 43205

**Curtis Okamoto,** Department of Pharmaceutical Sciences, University of Southern California, 1985 Zonal Avenue, Los Angeles, California 90033

**Richard M. Peek Jr.,** Division of Gastroenterology, Departments of Medicine and Cancer Biology, Vanderbilt University Medical Center, Room 1012A MRB IV, 2215 Garland Avenue, Nashville, Tennessee 37232

**Maikel P. Peppelenbosch,** Department of Cell Biology, University of Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands

**Mary H. Perdue,** Intestinal Disease Research Program, HSC-3N5C, McMaster University, 1200 Main Street West, Hamilton, Ontario, L8N3Z5, Canada

**J. Wesley Pike,** Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin 53706

**D. Brent Polk,** Departments of Cell and Developmental Biology and Pediatric Gastroenterology, Hepatology, and Nutrition, Digestive Disease Research Center, S4322 MCN Vanderbilt University, Nashville, Tennessee 37232-2576

**Barry C. Powell,** Child Health Research Institute, North Adelaide, and School of Pharmacy and Medical Sciences, University of South Australia; Department of Pediatrics, University of Adelaide, 72 King William Road, South Australia 5006, Australia

**María J. Pozo,** Department of Physiology, Nursing School, University of Extremadura, Avda Universidad s/n, 10071 Cáceres, Spain

**K. Ramaswamy,** Department of Medicine, University of Illinois at Chicago College of Medicine, 840 South Wood Street (m/c 716), Chicago, Illinois 60612

**Elyanne M. Ratcliffe,** Division of Gastroenterology, Hepatology, and Nutrition, Department of Pediatrics, Columbia University, 3959 Broadway, CHN 702, New York, New York 10032

**Drucilla J. Roberts,** Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts 02114

**Richard C. Rogers,** Department of Neuroscience, Pennington Biomedical Research Center, 6400 Perkins Road, Baton Rouge, Louisiana 70808

**Weifang Rong,** Department of Biomedical Science, University of Sheffield, Sheffield S10 2TN, United Kingdom

**Anil K. Rustgi,** Division of Gastroenterology, University of Pennsylvania School of Medicine, 415 Curie Boulevard, 600 Clinical Research Building, Philadelphia, Pennsylvania 19104

**Gino T. P. Saccone,** Department of General and Digestive Surgery, Flinders Medical Centre, Flinders Drive, Bedford Park, South Australia 5042, Australia

**Hamid M. Said,** Departments of Medicine and Physiology/Biophysics, University of California, Irvine, California, and VA Medical Center-151, Long Beach, California 90822

**Linda C. Samuelson,** Department of Molecular and Integrative Physiology, The University of Michigan, Ann Arbor, Michigan 48109-0622

**Guy R. Sander,** Child Health Research Institute, North Adelaide, and Department of Pediatrics, University of Adelaide, 72 King William Road, South Australia 5006, Australia

**Kenton M. Sanders,** Department of Physiology and Cell Biology, University of Nevada School of Medicine, Reno, Nevada 89557

**Sushil K. Sarna,** Division of Gastroenterology, Department of Internal Medicine, The University of Texas Medical Branch at Galveston, 9.138C Medical Research Building, 301 University Boulevard, Galveston, Texas 77555-1064

**Bellur Seetharam,** Departments of Medicine and Biochemistry, Medical College of Wisconsin and Clement Zablocki VA Medical Center, Research 151, 5000 National Avenue, Milwaukee, Wisconsin 53295

**Reza Shaker,** Division of Gastroenterology and Hepatology, Digestive Disease Center, Medical College of Wisconsin, 9200 West Wisconsin Avenue, Milwaukee, Wisconsin 53226

**Nirupama K. Shevde,** Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin 53706

**Xuan-Zheng Shi,** Division of Gastroenterology, Department of Internal Medicine, The University of Texas Medical Branch at Galveston, 9.138E Medical Research Building, 301 University Boulevard, Galveston, Texas 77555-1064

**Benjamin L. Shneider,** Department of Pediatrics, Box 1656, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, New York 10029

**Arthur Shulkes,** Department of Surgery, University of Melbourne, Austin Health, Heidelberg, Victoria 3084, Australia

**Johan D. Söderholm,** Department of Biomedicine and Surgery, Linköping University Hospital, Linköping SE-581 85, Sweden

**Ronald J. Sokol,** Department of Pediatrics, Section of Pediatric Gastroenterology, Hepatology, and Nutrition, Pediatric General Clinical Research Center, University of Colorado School of Medicine, The Children's Hospital, 1056 East 19th Avenue, Box B290, Denver, Colorado 80218

**Martin C. Steward,** Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, United Kingdom

**Judith Storch,** Department of Nutritional Sciences, Rutgers University, 96 Lipman Drive, New Brunswick, New Jersey 08901-8525

**Joseph H. Szurszewski,** Department of Physiology and Biomedical Engineering, Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, 200 First Street, SW, Rochester, Minnesota 55905

**Yvette Taché,** CURE/Digestive Diseases Research Center and Center for Neurovisceral Sciences and Women's Health, Department of Medicine, Division of Digestive Diseases, David Geffen School of Medicine at the University of California, Los Angeles, and VA Greater Los Angeles Healthcare System, CURE Building 115, Room 117, 11301 Wilshire Boulevard, Los Angeles, California 90073

**Jan Tack,** Department of Gastroenterology, University Hospitals Leuven, Center for Gastroenterological Research, University of Leuven, Herestraat 49, 3000 Leuven, Belgium

**Koji Takeuchi,** Department of Pharmacology and Experimental Therapeutics, Kyoto Pharmaceutical University, Misasagi, Yamashina, Kyoto 607, Japan

**Jay R. Thiagarajah,** Departments of Medicine and Physiology, University of California San Francisco, 1246 Health Sciences East Tower, San Francisco, California 94143-0521

**Maret G. Traber,** Department of Nutrition and Exercise Sciences, Linus Pauling Institute, Oregon State University, 571 Weniger Hall, Corvallis, Oregon 97331-6512

**R. Alberto Travagli,** Department of Neuroscience, Pennington Biomedical Research Center, 6400 Perkins Road, Baton Rouge, Louisiana 70808

**Eric Turk,** Department of Physiology, The David Geffen School of Medicine at University of California Los Angeles, Los Angeles, California 90095-1751

**Gijs R. van den Brink,** Laboratory for Experimental Internal Medicine, Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

**A. S. Verkman,** Departments of Medicine and Physiology, University of California San Francisco, 1246 Health Sciences East Tower, San Francisco, California 94143-0521

**Alda Vidrich,** Digestive Health Center of Excellence, University of Virginia, Charlottesville, Virginia 22908-0708

**V. K. Viswanathan,** Section of Digestive Diseases and Nutrition, University of Illinois, 840 South Wood Street, Room 738A (m/c 716), Chicago, Illinois 60612

**David B. Vodusek,** Division of Neurology, University Medical Center, Ljubljana 1525, Slovenia

**Sean M. Ward,** Department of Physiology and Cell Biology, University of Nevada School of Medicine, Reno, Nevada 89557

**Makoto Watanuki,** Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin 53706

**John A. Williams,** Department of Molecular and Integrative Physiology, University of Michigan Medical School, Ann Arbor, Michigan 48109-0522

**Allan W. Wolkoff,** Marion Bessin Liver Research Center and Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461

**Jackie D. Wood,** Departments of Physiology and Biology and Internal Medicine, The Ohio State University College of Medicine, 304 Hamilton Hall, 1645 Neil Avenue, Columbus, Ohio 43210-1218

**Ernest M. Wright,** Department of Physiology, The David Geffen School of Medicine at University of California Los Angeles, 10833 Le Conte Avenue, 53-263 Center for Health Sciences, Los Angeles, California 90095-1751

**Hua Xu,** Department of Pediatrics, Steele Children's Research Center, University of Arizona Health Sciences Center, 1501 North Campbell Avenue, Tucson, Arizona 85724-5073

**Elizabeth H. Yen,** Harvard Medical School Fellowship in Pediatric Gastroenterology and Nutrition, Gastrointestinal Cell Biology, Enders 720, Children's Hospital Boston, 300 Longwood Avenue, Boston, Massachusetts 02115

**Masaru Yoshida,** Frontier Medical Science in Gastroenterology, International Center for Medical Research and Treatment, Kobe University School of Medicine, 7-5-1 Kusunoki-cho, Chuo-ku, Kobe 650-0017, Japan

**David I. Yule,** Departments of Pharmacology and Physiology, University of Rochester Medical School, Rochester, New York 14642

**This Page Intentionally Left Blank**

# Preface to the First Edition

As with any publishing venture and especially one of this magnitude, one must first ask, "Why?" The Associate Editors and I were motivated primarily to collect in one set of volumes the most up-to-date and comprehensive knowledge in our field. Nothing comparable has been attempted in the area of gastrointestinal physiology during the past fourteen years. During this time, there has been a rapid expansion of knowledge and many new areas of investigation have been initiated.

More than fifty leading scientists—physiologists, clinical specialists, morphologists, pharmacologists, immunologists, and biochemists—have contributed chapters on their various areas of expertise for these volumes. Our original goal was to review the entire field of gastrointestinal physiology in one work. After examining all of the chapters, however, it was apparent that the final product encompassed more than physiology. The chapters reflect the backgrounds of the authors and the approaches of their different disciplines. As such, these volumes contain information for not only the investigator working in these fields but for the clinician or graduate student interested in the function of the gastrointestinal tract. Anyone involved in teaching gastrointestinal physiology of pathophysiology can readily find the latest and most pertinent information on any area in the discipline.

This work is divided into five sections. The first consists of topics such as growth, the enteric nervous system, and gastrointestinal peptides, each of which relates to all areas of the gastrointestinal tract. The second section contains material describing smooth muscle physiology and gastrointestinal motility. The third section presents treatment of the functions of the stomach and pancreas. The fourth series of chapters treats the entire field of digestion and absorption. These chapters vary from basic electrophysiology and membrane transport to reviews of mechanisms leading to clinical conditions of malabsorption. The final section contains chapters on areas peripheral to physiology (such as immunology, parasitology, and prostaglandins) yet necessary for a comprehensive understanding of the subject.

No one person can presume to organize and edit a scientific work of this scope. I was fortunate to enlist the aid of four preeminent scientists whose expertises cover the entire field. James Christensen was primarily responsible for the chapters on smooth muscle and motility. Eugene D. Jacobson solicited and edited most of the chapters dealing with secretory mechanisms as well as those covering many of the general topics. Chapters relating to regulation were primarily handled by Morton I. Grossman, and those covering aspects of digestion and absorption were organized and reviewed by Stanley G. Schultz. I am exceedingly grateful to these four men without whom this work would not have been possible.

L.R.J.

**This Page Intentionally Left Blank**

# Preface

This fourth edition of *Physiology of the Gastrointestinal Tract* follows 12 years after the third edition. The delay was mainly due to buyouts and mergers of the involved publishing houses, certainly not to a lack of new information. On the contrary, the explosion of information at the cellular level, made possible, in part, by the continued emergence of powerful molecular and cellular techniques, has resulted in a greater degree of revision than that of any other edition. Section I, now titled "Basic Cell Physiology and Growth of the Gastrointestinal Tract," contains numerous new chapters on topics such as transcriptional regulation, signaling networks in development, apoptosis, and mechanisms in malignancies. Most of the chapters in the first section have been edited by Juanita L. Merchant. Section II has been renamed "Neural Gastroenterology and Motility" and has been expanded from 7 chapters with rather classic titles to more than 20 chapters encompassing not only the movement of the various parts of the digestive tract but also cell physiology, neural regulation, stress, and the regulation of food intake. Almost all of the chapters in the second section have been recruited and edited by Jackie D. Wood. Section III is entirely new and contains chapters on "Gastrointestinal Immunology and Inflammation," which were edited by Kim E. Barrett. Section IV, "Physiology of Secretion," consists of chapters with familiar titles but with completely updated information to reflect the advances in our understanding of the cellular processes involved in secretion. Section V, "Digestion and Absorption," contains new chapters on the intestinal barrier, protein sorting, and ion channels, together with those focusing on the uptake of specific nutrients. These chapters have been recruited and edited by Hamid M. Said and Fayez K. Ghishan.

The original purpose of the first edition of this textbook—to collect in one set of volumes the most current and comprehensive knowledge in our field—was also the driving force for this edition. As mentioned earlier, this edition includes completely new chapters that cover many new areas. Although the number of chapters has increased by 15, some chapters from the previous edition have been eliminated, some with identical titles have been written by different authors, and a few have been updated by the original authors. The final product again encompasses more than physiology. The information provided is relevant not only to the researcher in the various specialized areas but also to the clinical gastroenterologist, the teacher, and the student. The authors have done an excellent job of presenting their knowledge in a style that is readable and understandable.

Much of the effort in organizing and editing these volumes has come from five preeminent scientists whose interest and expertise cover the entire field. Drs. Barrett, Ghishan, Merchant, Said, and Wood met with me to decide on chapter topics, authors, and the overall organization of the material. They were responsible for recruiting authors and for the scientific editing of most chapters. The enthusiasm and abilities of these individuals simplified my task as editor, and without them this work would not have been possible. I also am especially grateful to Philip Carpenter of Elsevier, who contacted authors, tracked submissions, and assisted me in many ways.

My Associate Editors and I are all grateful to the contributing authors who were generous enough to make their expert knowledge available. Their efforts have made this work more than a mere review of past contributions to a field. The various chapters synthesize and criticize this accumulated knowledge and identify voids in it, pointing out future directions for research; many of them are superb presentations of information in fields that have been reviewed nowhere else.

L.R.J.

# Acknowledgments

From the organization stage to actual production the following people provided invaluable assistance, helpful suggestions, and a great deal of support. Their role and efforts have been much more than what is normally provided by a publisher, and I express my thanks to them.
Jasna Markovac (Senior Vice President, Global Academic & Customer Relations)
Julie Eddy (Publishing Services Manager)
Lisa Royse (Production Editor)
Mara Conner (Publishing Editor)
Judy Meyer (Publishing Editor)
Tari Broderick (Senior Publishing Editor)
Cate Rickard Barr (Design Manager)
Andrea Lutes (Book Designer)
Patricia Howard (Senior Marketing Manager)
Trevor Daul (Senior Marketing Manager)
Philip Carpenter (Developmental Editor)

CHAPTER 1

# Transcriptional and Epigenetic Regulation

Juanita L. Merchant and Longchuan Bai



With the completion of the human genome sequencing project at the dawn of the third modern millennium, we have come to appreciate that we are only at the start of a new era of genomic enlightenment. Perhaps the most important piece of information that we have learned is that the clues to our genetic destiny are contained in more than just the primary sequence of DNA. Apparently, what distinguishes humans from other life-forms, and most interestingly, other mammals, lies in the complex modifications and function of the 20,000 to 30,000 genes. Not only are these 25,000 or so genes alternatively spliced, but these products are chemically modified to change their function. Therefore, as opposed to our genetic template being composed of a mere 25,000 genetic units, we are actually controlled by 25,000 to the $n^{th}$ power. The latter value has yet to be determined, but likely results in an enormous combination of genetic events. This chapter reviews what has led us to reformulate our notions of gene expression in the postgenomic era.

J. L. Merchant: Departments of Internal Medicine and Molecular and Integrative Physiology, University of Michigan, Ann Arbor, Michigan 48109.

L. Bai: Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan 48109.

*Physiology of the Gastrointestinal Tract, Fourth Edition,* edited by Leonard R. Johnson. Academic Press, 2006.

## OVERVIEW OF GENE ORGANIZATION

### Gene Composition

The molecular definition of a eukaryotic gene is complex, but in the simplest terms, it is a nucleic acid sequence that encodes one polypeptide or messenger ribonucleic acid (mRNA) molecule (1). Genes are composed of two intertwining polymers of DNA that are noncovalently attached to a variety of proteins, including histones and specialized proteins (e.g., polymerases and various accessory proteins). The association of DNA, histones, and specialized nuclear proteins collectively is called chromatin. Chromosomes are composed of continuous strands of chromatin that have been compacted by supercoiling and looping to fit into the nucleus. Most importantly, they are the basic heritable unit in the mammalian cell. In humans, there are 46 chromosomes, or 23 pairs. The smallest unit of the DNA polymer is a nucleotide, a base attached to the first carbon of a five-carbon sugar phosphorylated at its fifth carbon (Fig. 1-1). Nucleosides do not contain phosphates; thus, they differ from nucleotides, which contain one, two, or three

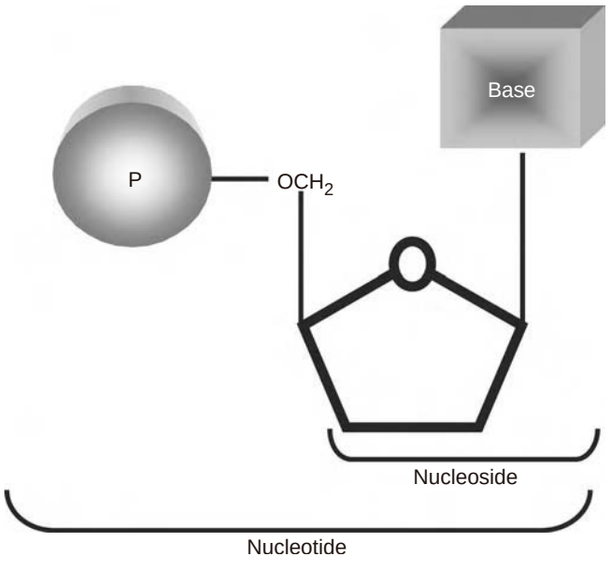



**FIG. 1-1.** Nucleic acid structure. A nucleoside consists of a purine or pyrimidine base covalently linked to the first carbon of the pentose ring. The addition of one, two, or three phosphate groups yields a nucleotide monophosphate, diphosphate, or triphosphate, respectively. The type of sugar determines the type of nucleic acid: ribose in ribonucleic acid (RNA) and deoxyribose in DNA.

phosphate groups. The four nucleotides are distinguished by the type of base that they contain: adenine (A), thymine (T), cytosine (C), or guanine (G). DNA contains the sugar *deoxy*ribose, whereas RNA contains the sugar ribose and the base uracil (U) instead of thymine.

Polymers of nucleotides or nucleic acids (also called nucleoside monophosphates, diphosphates, or triphosphates) are formed when the free phosphate group attached to the fifth carbon of an adjacent nucleotide of the pentose sugar condenses with the hydroxyl group on the third pentose carbon to produce two ester bonds and water (phosphodiester bond). Accordingly, the proximal end of each DNA strand (5′ end) contains a phosphate group in the 5 position of the deoxyribose sugar residue. The terminal nucleic acid at the 3′ end of each DNA strand contains a free hydroxyl group in the 3 position of the deoxyribose ring. By convention, nucleotide sequences are written from 5′ to 3′, reading from left to right, with the sense strand presented as the upper strand. The antisense strand, written on the bottom, is antiparallel and complementary to the sense strand so that the 5′ to 3′ direction proceeds from right to left. Each nucleotide within the polymer is base paired with a particular nucleotide on the opposing strand by hydrogen bonds; adenine pairs with thymine, and guanine pairs with cytosine. The DNA strand containing the same sequence as the mRNA is designated the *sense strand,* and the strand that it pairs with is designated the *antisense strand.* The antisense strand becomes the template sequence that will be transcribed by RNA polymerase II (Pol II) into mRNA and subsequently translated into amino acids.

Most studies on transcriptional control focus on genes transcribed by the seven-subunit enzyme Pol II, and thus are designated as class II genes (2). It is Pol II that is responsible for transcribing gene sequences into protein-encoding mRNA. Only 4% of the total RNA in the cell is mRNA. Many of these initial primary transcripts (heterogeneous nuclear RNA [hnRNA]) are further processed as discussed later. Nine percent of cellular RNA is hnRNA, the bulk of which are small nuclear RNA (snRNA; e.g., U2 involved in RNA splicing, 4%) and small nucleolar RNA (e.g., U22 snoRNA comprising 1%). The other 4% of hnRNA is mRNA. An additional 1% of total cell RNA is called guide RNA, which edits mature mRNA transcripts (3). RNA polymerase I (Pol I) transcribes all of the ribosomal genes except for the *5S* gene. Ribosomal RNA represents about 75% of the RNA in the cell. RNA polymerase III (Pol III) transcribes the *5S* ribosomal gene and the genes encoding transfer RNA. Transfer RNA represents about 15% of the total RNA in the cell. Pol I and III transcribe genes that will not be further translated into peptides, although their primary transcripts are also processed before reaching the cytoplasm. Because Pol II transcribes genes encoding proteins and peptides, Pol II–regulated genes are the primary focus of this chapter.

One may conceive of a gene as being analogous to a long sentence read from left to right and composed of letters organized into words separated by spaces and marks of punctuation. Specific DNA sequences "punctuate" the gene with important start and stop signals for transcription and translation. One gene may comprise several hundred to several thousand DNA base pairs. These base pairs (the alphabet) are organized into functional groups (phrases) based on whether a particular sequence is untranscribed, only transcribed, or both transcribed and translated (Fig. 1-2). Exons are DNA sequences that are transcribed into mRNA by Pol II and *ex*it the nucleus. Within the cytoplasm, exons may or may not be translated into peptides. Those exons that are transcribed *and* translated form the coding sequences (coding exon). In general, the term *intron* is used to describe the *in*tervening DNA sequence that is transcribed but is removed from the primary transcript by RNA splicing (RNA processing) before it exits the nucleus as a mature transcript (see Posttranscriptional Processing later in this chapter and also Chapter 2). DNA sequences or elements that regulate transcription and are not transcribed into mRNA usually reside in the 5′ portion of a gene upstream (to the left of) of the promoter. The promoter is a group of DNA sequences that binds Pol II in concert with accessory proteins to initiate the synthesis of mRNA. Accessory proteins control the accuracy and rate of polymerase binding. The first nucleotide transcribed into mRNA is assigned the number *1* with subsequent nucleotides (downstream or to the right of the promoter) assigned positive numbers as transcription proceeds toward the 3′ end. Nucleotides preceding the promoter (upstream or 5′) are assigned negative numbers. DNA sequences that encode a polypeptide (open reading frame) begin with the translational start site codon ATG (encoding methionine) and end with one of the three stop codons: TAA, TAG, or TGA.

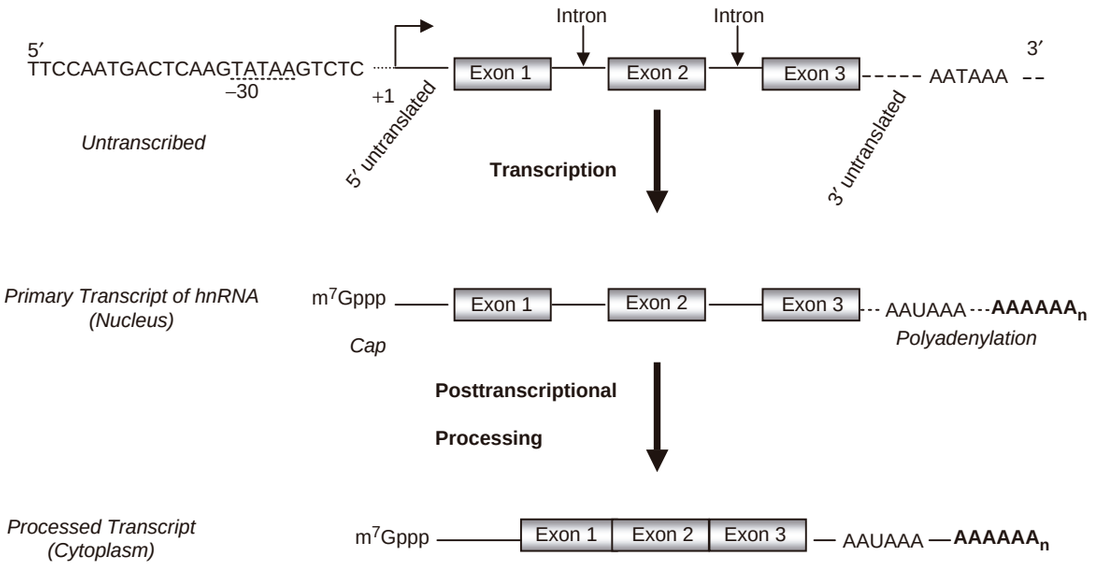



**FIG. 1-2.** Gene structure, transcription, and posttranscriptional processing. A gene is composed of several hundred to several thousand base pairs, subdivided into functional elements. The location of 5′ and 3′ untranslated sequences, exons, and introns are shown. The 5′ flanking sequences contain specific DNA elements (e.g., TATA box). Ribonucleic acid (RNA) polymerase II transcribes DNA into heterogeneous nuclear RNA (hnRNA) during *transcription.* Twenty base pairs after the sequence AATAAA is transcribed to AAUAAA, messenger RNA (mRNA) are cleaved and the polyadenylate tail is added to the 3′ end. A methylated guanylate residue is added to the 5′ end of the mRNA through a triphosphate linkage. Before exiting the nucleus, intron segments are removed by splicing factors during *posttranscriptional processing.*

(The translational start and stop codons, respectively, are transcribed into mRNA as AUG, UAA, UAG, and UGA.) Because one amino acid is encoded by three nucleotides or a triplet (codon), two or three peptides may be encoded by overlapping codons simply by shifting the reading frame by one or two nucleotides. Regulatory sequences that are transcribed but not translated reside at both the 5′ and 3′ ends of the mature RNA transcript. Both 5′ and 3′ untranslated regulatory sequences, which range from 10 to several thousand nucleotides, are thought to participate in the fidelity of translation and mRNA stabilization or destabilization.

RNA molecules that encode proteins (except most histone proteins) are distinguished from ribosomal and transfer RNA by the series of adenosines added to the 3′ end of the molecule (poly(A) RNA; see Fig. 1-2). This feature is a useful means to isolate mRNA from other, more abundant RNA species (transfer and ribosomal RNA) and also designates the functional termination of the protein-encoding portion of the gene. During transcription, the primary RNA transcript is cleaved 20 bp downstream of the AAUAAA site at the 3′ end, and ~150 to 200 adenine nucleotides are added to form the poly(A) tail (4–6). The 5′ end of the mRNA transcript receives a protective "cap" after synthesis of the first 30 nucleotides, which consists of a guanylate residue methylated at the 7 position and linked to the first nucleotide of RNA by three phosphates. The RNA cap is a high-affinity binding site for ribosomes (7,8). Notably, the element AATAA that signals the site of the poly(A) tail is not necessarily the functional end of the gene. Rather, the 3′ untranslated region (3′UTR) and 3′ untranscribed regions may still contain regulatory elements that can modulate gene expression. Therefore, just as the 5′ end of a gene must be determined empirically, so must the 3′ end of the gene.

The 5′ border of a gene is identified by the promoter region (functionally determined) and structurally by the first nucleotide transcribed into mRNA (cap site) as determined by various reverse transcriptase methods—for example, primer extension analysis or anchored polymerase chain reaction (PCR) (9). These techniques use reverse transcriptase to synthesize complementary or copy DNA (cDNA; Fig. 1-3). Radiolabeled primers complementary to the 5′ end of the DNA sequence to be copied are allowed to anneal to mRNA. Reverse transcriptase then adds deoxynucleotides to the primer in the 3′ to 5′ direction. Synthesis of the cDNA will terminate when the 5′ end of the mRNA is reached. Template mRNA molecules are removed by ribonucleases (RNases), and the synthesis of a double-stranded cDNA is completed through the action of DNA polymerase. Because the newly synthesized cDNA is radiolabeled at the 5′ end, the length of the cDNA (and hence the transcriptional start site) is determined by resolving the fragments on a denaturing polyacrylamide gel and comparing the length observed in base pairs to the known cDNA sequence. cDNA is also a useful tool for making probes to detect complementary nucleotide

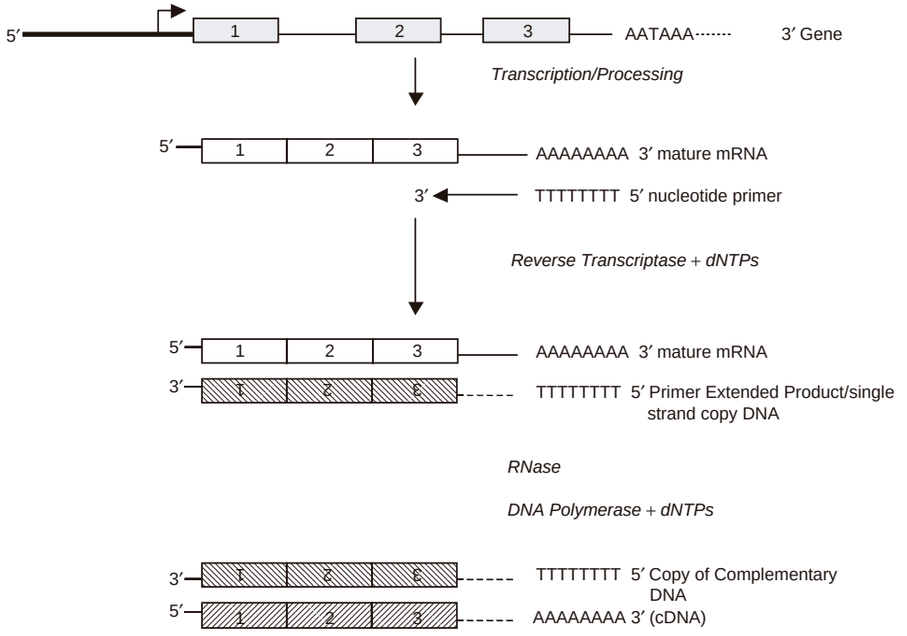



**FIG. 1-3.** Complementary DNA (cDNA). Primers complementary to a portion of the messenger ribonucleic acid (mRNA) are allowed to anneal. For unknown sequences, as in the synthesis of cDNA libraries, a primer complementary to the poly(A) tail is used, that is, poly (dT). Reverse transcriptase added together with all four deoxynucleotides (dNTPs) will transcribe mRNA in the 3′ to 5′ direction to make cDNA. The mRNA template is removed by RNases, and double-stranded cDNA is made using DNA polymerase. In primer extension analysis, the 5′ end of mRNA (the cap site) is identified by annealing primers of a known sequence near the 5′ end of mRNA.

sequences and for making cDNA libraries that reflect the spectrum and relative abundance of specific mRNA within a given cell. These cDNA libraries must be contrasted with genomic phage libraries in which the DNA sequences in the phage heads reflect the number of times that a particular gene sequence is represented in the host genome, which is usually once.

The 5′ sequences flanking the gene are defined functionally by various methods other than simple structural information. These sequences direct the developmental, tissue-specific, and inducible expression of the gene and can range from a few hundred to several thousand base pairs (10). It is possible to identify the sequences conferring these regulated gene activities by using methods such as DNA transfer into cell lines (11,12) and transgenic mouse models (13,14). For example, the expression of gastrin in the adult occurs in the antrum of the stomach and in the first portion of the duodenum (15–17). However, gastrin is never expressed in skin or kidney. Thus, if 1000 bp of 5′ flanking sequence permits the expression of gastrin in a fibroblast or kidney cell line, but 20,000 bp do not, it may be concluded that the untranscribed sequences between −1000 and −20,000 bp from the promoter are important in shutting off expression of gastrin in skin and kidney, sites where gastrin is never expressed *in vivo.* Thus, the 5′ regulatory sequences important in normal expression of the gastrin gene may extend as far upstream as −20,000 bp from the start site of transcription. Alternatively, the 5′ or even 3′ borders may extend even further if functional data indicate that a larger sequence is required for the appropriate tissue and temporal expression to be observed with the native gene. Recently, it has been found that there are specific DNA elements called Insulator elements that mark the boundary of genes (18). These elements, originally identified on the globin gene, bind a transcription factor called CTCF and are capable of preventing the spread of histone acetylation between adjacent genes (19). Specific examples of tissue-specific elements have been reported within the promoters of several gastrointestinal (GI) peptides (e.g., gastrin and secretin), as well as for specific intestinal proteins (e.g., sucrase-isomaltase) (20–26).

Similar experiments may also be performed in transgenic mice with constructs containing various lengths of 5′ flanking sequences regulating reporter gene expression. Instead of transferring these reporter constructs into cell lines, they are injected into fertilized eggs and reimplanted into ovulating female mice to be expressed in the mouse germ line (13,14). The expression of these constructs in the offspring is analyzed by cytochemical detection of reporter gene products in various organs or in response to physiologic inducers (27). The transgenic approach to gene expression, like the experiments described earlier, permits anatomic, environmental, and developmental analysis in the whole animal (28,29). This approach is particularly valuable in understanding the

regulatory sequences important for the tissue-specific expression of the genes in different cell types of the same organ such as the small intestine (30,31). Given the requirement for larger and larger pieces of DNA to recapitulate native expression in transgenic mouse models, techniques have been developed to clone and manipulate large pieces of DNA (more than 50 kilobases; e.g., yeast artificial chromosomes [YACs] and bacterial artificial chromosomes [BACs]) (32,33). Recombineering is a powerful technique performed in bacteria that permits introduction of foreign DNA or point mutations into these large plasmids that are eventually introduced into transgenic mice (34–37).

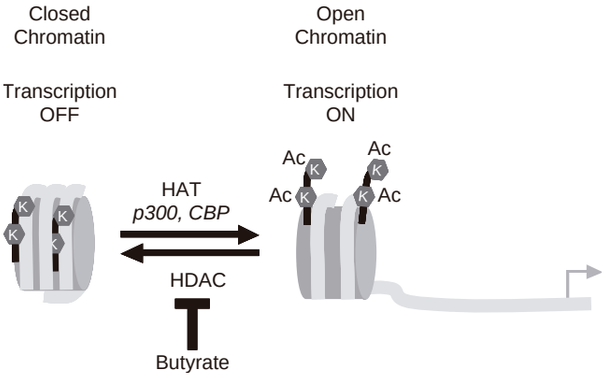



**FIG. 1-4.** Nucleosome structure. The double-strand DNA helix winds twice around a complex of the four core histones assembled as dimers. Unacetylated histones are positively charged and adhere tightly to the negatively charged DNA, preventing access by transcription regulatory proteins. Histones that are acetylated are less positively charged and do not adhere as tightly to chromatin, allowing access of regulatory proteins to the DNA. The addition or removal of acetyl groups to the ends of histones is regulated by histone acetyltransferase (HATs) and histone deacetylase enzyme complexes (HDACs). The short-chain fatty acid butyrate inhibits the activity of HDACs.

## EPIGENETIC INFLUENCES

*Epigenetics* literally means "outside of or beyond genetics," and it refers to the "study of genetic modifications that are mitotically and/or meiotically heritable yet do not change the DNA sequence" (38). Thus, mutations or deletions alter the character or length of the sequence that, in turn, alters the primary sequence of the protein. By contrast, epigenetic influences chemically modify the nucleotide or amino acid structure that, in turn, changes how that particular residue is recognized by nuclear proteins, without changing the sequence itself. Although it is now clear from the completed sequence of the human genome that there are only about 20,000 to 30,000 gene loci, the complexity of the genetic information encoded in human chromosomes must enlist other features of chromatin (39). The epigenetic influences on chromatin appear to be one of the critical features that enhance genomic complexity. Major targets of epigenetic changes are histones, basic proteins coating the naked DNA double helix. The N-terminal tails of histones (H1, H2A, H2B, H3, H4) are positively charged because of the basic amino acid lysine. The positively charged histones attach to DNA because of the negatively charged phosphate backbone of DNA. The ionic interaction is reduced if the positive charge on the lysines is removed. Specific enzymes called histone acetyltransferases (HATs) acetylate the lysine side group, effectively eliminating the positive charge (Fig. 1-4). The loss of the ionic interaction between the histones and phosphate groups on DNA permit greater access to the DNA helix by accessory proteins such as polymerases, transcription factors, and coactivators or repressors. DNA in the form of chromatin becomes open, accessible, and readily transcribed. By contrast, there are enzymes that will "close" chromatin by removing the acetyl groups from the lysines at the N-terminal tails of histone proteins. These enzymes are called histone deacetylases (HDACs). Removal of the acetyl group restores the positive charge to the histones allowing the ionic interaction between histones and DNA to be restored. The nonhistone proteins such as polymerases and transcription factors become excluded from DNA, transcription is silenced, and chromatin is inactive.

Collectively, the histones and accessory proteins associated noncovalently with DNA are what forms chromatin. Chromatin exists in two forms: euchromatin and heterochromatin (40). Euchromatin contains the actively transcribed genes and becomes decondensed during DNA replication. Euchromatin is also centrally located in the nucleus. By contrast, heterochromatin contains transcriptionally silent genes that remain condensed at the periphery of the nucleus. The DNA sequences within heterochromatin are repetitive, and only 15% of nuclear chromatin is heterochromatin. The major forms of epigenetic modifications in mammalian cells occur on DNA and histones and include such covalent modifications as acetylation and methylation, but also via the addition of other organic residues. These epigenetic changes affect such events as chromatin folding, gene expression, X-chromosome inactivation, and genomic imprinting (41). Epigenetic events are essential for development and differentiation, during which clusters of genes must be activated or silenced at precisely timed intervals to allow for the organism's growth and maturation.

### Histone Modifications

The basic repeating unit of chromatin is the nucleosome. Each nucleosome is composed of 147 bp of DNA wrapped twice around a histone protein octamer consisting of 2 molecules of each of the 4 core histones (H2A, H2B, H3, and H4). The linker histone H1 sits alone between each core nucleosome, facilitating further compaction (42). Each histone contains a structured globular domain with a histone-fold motif important for nucleosome assembly and a highly charged unstructured amino-terminal tail of 25 to 40 residues,

which protrudes from the body of the nucleosome to latch onto the phosphate backbone. The amino termini are the major sites for histone modifications (43). Histones can be modified by acetylation, methylation, phosphorylation, adenosine diphosphate (ADP)-ribosylation, ubiquitination, and sumoylation (44). The mixture of these covalent modifications creates a "code" on the surface of the histone molecule that is subsequently recognized by bromo and chromo domain–containing proteins mediating chromatin compaction, transcription, and DNA repair (45). Acetylation, methylation, ubiquitination, and sumoylation occur on the lysine residues, whereas methylation also occurs on arginine residues. Phosphorylation occurs on serines and threonines, and ADP-ribosylation occurs on glutamic acids. Most of these modifications, particularly acetylation, alter the charge distribution on the amino terminus and also alter nucleosome structure, which may, in turn, regulate chromatin structure (46,47). Some covalent modifications act as molecular switches, enabling or disabling subsequent covalent modifications, which explains the functional complexity of epigenetic modifications (48). Therefore, each modification correlates with a specific physical status of chromatin.

### *Histone Acetylation*

Acetylation of histones occurs at the ε-amino side group of specific lysines within the N termini of histones. HATs transfer an acetyl group from acetyl-coenzyme A as a donor to the histone terminal lysines (49). In hypoacetylated chromatin, the positive charges on unacetylated lysines are attracted to the negatively charged DNA, producing compact, closed chromatin thereby repressing transcription (50). In contrast, acetylation of the lysines removes their positive charges, resulting in a less compact, open chromatin structure, which facilitates gene transcription. Therefore, HAT activity, and subsequently histone acetylation, is linked mainly to transcriptional activation (51) (see Fig. 1-4). Removal of the acetyl group (deacetylation) by HDACs restores the positive charge on lysines, and chromatin becomes compacted and less accessible to regulatory proteins required for transcription. Thus, HDACs and deacetylation are primarily associated with transcriptional repression (see Fig. 1-4).

HATs are divided into five families. These include the p300/CBP (cyclic 3′,5′-adenosine monophosphate [cAMP] response element binding [CREB] protein) HATs (p300 and CBP), Gcn5-related acetyltransferases (GNATs; including Gcn5, p300/CBP-associated factor [PCAF], etc.), MOZ, Ybf2, Sas2, and Tip60 (MYST) (monocytic leukemia zinc finger protein [MOZ], Ybf2/Sas3, Sas2, and Tip60)-related HATs, the general transcription factor (GTF) HATs (TFIID subunit TAF250 and TFIIIC), and the nuclear hormone-related HATs (SRC1 and ACTR) (52). The most consistent functional characteristic of the HATs is that they are transcriptional coactivators. These proteins are components of large multisubunit complexes that do not bind DNA directly, but instead form protein–protein interactions with DNA-binding transcription factors (53).

The more numerous mammalian HDACs have been grouped into three protein classes (54). Class I includes HDACs 1, 2, 3, and 8. Class II includes HDACs 4, 5, 6, 7, 9, and 10. The class III HDAC family consists of the conserved nicotinamide adenine dinucleotide (NAD)–dependent Sir2 family of deacetylases. Like HATs, HDACs do not bind directly to DNA but rather are recruited by large multisubunit complexes to function primarily as corepressors of transcription (55).

The function of HATs and HDACs are of particular relevance in the GI tract because of the effect of butyrate, a by-product of colonic bacterial fermentation, on histone acetylation (see Fig. 1-4). Epidemiologic studies uniformly concur that a diet high in fiber is protective against colon cancer (56). The short-chain fatty acid butyrate is one of several fiber-derived fermentation products capable of maintaining epithelial cell differentiation (57). The differentiation effects were initially demonstrated after treatment of erythroleukemic cells with butyrate (58). Subsequently, it was discovered that the induction of differentiation by butyrate correlated with histone hyperacetylation (59–61) due to suppression of HDACs (62–66). Thus, the HDAC effects of butyrate and resulting histone hyperacetylation may, in fact, be one mechanism by which dietary fiber exerts its anticancer effects (67).

Reviews support the viewpoint that butyrate is a potent anticancer agent (68–70). Collectively, early studies emphasized the global effects of butyrate on chromatin remodeling, but the molecular basis for the gene-specific effects of butyrate remains poorly defined. HDAC inhibitors regulate less than 10% of actively transcribed genes. Most of those are up-regulated through GC-rich sites (71,72). In addition to histone acetylation, it is now known that DNA-binding proteins can become acetylated (52). Thus, a possible mechanism by which hyperacetylation induced by butyrate might target specific genes is through acetylation of specific transcription factors. The proposed function of acetylated transcription factors varies and includes increased or decreased DNA binding, as well as protein stability (73). In many instances, the genetic targets of butyrate are GC-rich sequences that bind Sp1 and Sp3. Gamma glutamyl transferase (74), insulin-like growth factor (IGF) binding protein 3 (75), G α(i2) (76), galectin (77), Cox1 (78), and intestinal alkaline phosphatase (79) are all up-regulated by butyrate through Sp1 sites. Sp1 binding sites are also implicated in the butyrate induction of $p21^{WAF1}$ gene expression (80). HAT p300, recruited to the $p21^{WAF1}$ promoter, cooperates with Sp1 and Sp3 to mediate the effects of butyrate (81). However, Sp1 does not cooperate directly with p300, but instead binds HDAC1 (82,83). The Sp1-HDAC1 complex, in turn, forms complexes with other corepressors such as Sin3A (84). Thus, Sp1 appears to be the factor that confers $p21^{WAF1}$ promoter repression by recruiting HDACs and corepressor complexes.

HDACs can have opposing functions, especially in cancer. HDACs can prevent the activation of tumor suppressor genes and block the ability of a cancer cell to undergo

apoptosis (85). However, HDAC2 silencing can trigger apoptosis (86). Another important feature of HDACs is their interaction with DNA methylation. HDACs cooperate with DNA methyltransferases (DNMTs) by removing the acetyl groups that would otherwise block methylation targets on histones or DNA (87,88).

### *Histone Methylation*

There are two types of histone methylation, targeting either lysine or arginine residues. Histone methyltransferases perform these modifications using S-adenosyl-methionine as the methyl group donor. Lysine methylation is implicated in changes in chromatin structure and gene regulation, whereas arginine methylation correlates with the active state of transcription, such as acetylation (89).

#### *Histone Methylation at Lysines*

Methylation of lysines residues (K) occurs on histone H3 primarily at K4, K9, and K27 and on H4 at K20 (Fig. 1-5). The lysine residues can be monomethylated, dimethylated, or trimethylated at the ε-amino group. The methylation of H3 is associated with an open chromatin configuration and gene activation (90,91). In contrast, the methylation of H3 at K9 is associated with condensed, repressed chromatin (92).

In general, there are at least four families of lysine methyltransferases. All of the lysine methyltransferases are distinguished by the presence of Su(var)3-9, Enhancer of Zeste, and Trithorax (SET) domains. The fourth family of these methyltransferases contains other protein domains aside from the SET domain. SET protein domains are approximately 130 residues homologous to amino acid segments in SET, three *Drosophila* proteins with intrinsic methyltransferase activity (93,94). The mammalian form of Su(var)3-9 is SUV39H and is involved in stabilizing heterochromatin by trimethylation of histone H3 at lysine K9. The trimethyl group creates an atomic feature or imprint on H3 that, in turn, is recognized by HP1, a chromatin organization modifier (chromo domain proteins) (95). The methylated or acetylated imprints on DNA or histones are recognized by two classes of proteins: those with chromo domains that recognize methyl group imprints and those with bromo domains that recognize acetyl group imprints. Transcriptional coactivators such as CBP, p300, and PCAF are HATs that contain bromo domains. They acetylate histones and other nuclear proteins; thus, not surprisingly, they also recognize an acetyl group imprint. These proteins are discussed in greater detail later in this chapter in Chromatin-Binding Proteins.

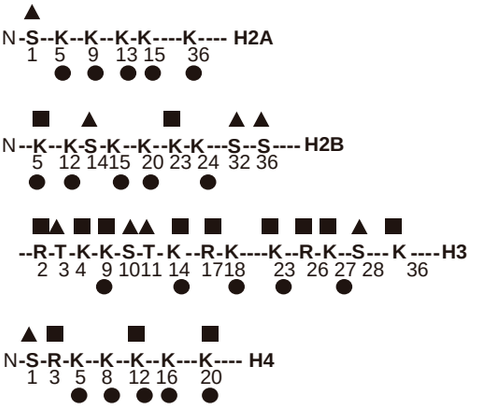



**FIG. 1-5.** Histone modifications on histone tails. Shown are the amino-terminal histone residues modified by acetylation (*filled circles*), methylation (*filled squares*), and phosphorylation (*filled triangles*).

#### *Histone Methylation at Arginines*

Methylation at arginines occurs within the tails of histones H3 (R2, R17, and R26) and H4 (R3) and is catalyzed by coactivator-associated arginine methyltransferase 1 (CARM1) and protein arginine *N*-methyltransferase 1 (PRMT1), respectively, in mammalian cells (see Fig. 1-5). Like lysines, arginines can be either monomethylated or dimethylated (asymmetric or symmetric) on the guanidino nitrogen, and this process is antagonized by human peptidylarginine deiminase 4 (PADI4), which converts methyl-Arg to citrulline (96,97). Less is known about the fate of histones methylated at arginines. However, initial studies indicate that the methylated arginines create an imprint recognized by coregulatory molecules, for example, p300 and switching/sucrose nonfermenting (SWI/SNF) (98,99).

### *Histone Phosphorylation*

Histone phosphorylation occurs on all four core histones: H2A (S1), H2B (S14), H3 (S10 and S28), and H4 (S1) (see Fig. 1-5). The phosphorylation of S10 in H3 is associated with transcriptional activation (100) and chromosome condensation during mitosis (101). In addition, phosphorylation of S10 in H3 is also associated with the transduction of external signals to chromatin, leading to the transient expression of immediate early genes (102,103). The phosphorylation of H3 is mediated by several specific kinases, activated by distinct pathways. For example, mammalian mitotic H3 phosphorylation is associated with Aurora B kinases (104,105), H3 phosphorylation by IKKα is important for the activation of nuclear factor (NF)-κB (106), and the immediate early gene response is mediated mainly by *m*itogen and *s*tress-activated *k*inases MSK1 and MSK2 (107). Histone H2B phosphorylation condenses the chromatin and is involved in apoptosis (108,109). The downstream effects of phosphorylation of H2A and H4 are unknown.

Of the histone modifications, acetylation and phosphorylation are reversible. Consequently, if the presence of a modification influences transcription in a particular way, its removal may have the opposing effect. In this way the cell could effectively respond to changes in environmental cues. Different histone modifications may be linked mechanistically. For example, phosphorylation of S10 on H3 enhances

histone acetylation by Gcn5 (110,111), whereas H3 K9 methylation inhibits phosphorylation at H3 S10 (93). Given the number of sites and the variety of possible modifications, the combinatorial possibilities are extremely large. The combinatorial pattern of N-terminal modifications results in a heterogeneous identity for each nucleosome that the cell interprets as a readable code from the genome to the cellular machinery directing various processes to occur. This concept is commonly referred to as the "histone code hypothesis" (45). The precise modification status of a given histone tail on a given gene can also change during the process of transcriptional regulation and each of these different constellations of histone modifications may elicit distinct downstream transcriptional signals (45).

### DNA Methylation

DNA methylation is a postsynthesis modification that normal DNA goes through after each replication. This modification is catalyzed by DNMTs and occurs on the C-5 position of cytosine residues within CpG dinucleotides located primarily in the promoter of a gene. There are three major DNMTs (DNMT1, DNMT3a, and DNMT3b). Each DNMT plays a distinct and critical role in cells. Murine knockouts of DNMT1 and DNMT3b exhibit embryonic lethality (112). The DNMT3a homozygous knockout mouse appeared normal at birth but died by aged 4 weeks (112, 113). In humans, mutations of DNMT3b are linked to ICF syndrome (*i*mmunodeficiency, *c*entromere instability, *f*acial anomalies) (112,114). Sixty percent of human genes contain a CpG island (115). Although methylation can also occur in other parts of the gene, CpG dinucleotides tend to be underrepresented in the genome, and when they are found, they appear in clusters ranging from 0.5 to several kilobases with a GC content greater than 55% (116). These clusters are known as CpG islands (117). Methylation of CpG islands is a late evolutionary development and functions to maintain genome stability by repressing transposons and repetitive DNA elements (118).

DNA methylation is an important player in many processes, including transcriptional repression, X-chromosome inactivation, and genomic imprinting. CpG islands located in the promoter region of genes are normally hypomethylated about 40% of the time (116). Their hypermethylation causes stable heritable transcriptional silencing. As observed with HDACs and deacetylation, the methylation status in cancers may seem contradictory. Aberrant *de novo* hypermethylation of CpG islands is a hallmark of some human cancers and is found early during carcinogenesis (119–121). Tumor suppressor genes are locally hypermethylated in some cancers to silence their expression, whereas oncogenes may be hypomethylated (116). Tumor cells globally demonstrate an overall hypomethylation of DNA, a process that has more recently been linked to nutrition (122). S-adenosylmethionine is the primary methyl donor in the cell and is reduced in conditions predisposed to cancer (123).

Genomic imprinting occurs in gametogenesis and is necessary for development. One of the X chromosomes in female individuals is not expressed because of the heavy methylation of the inactive X chromosome. The epigenetic phenomenon whereby expression of a gene depends on whether it is inherited from the mother or the father is called imprinting, and is caused by differential methylation of specific cytosine bases on the maternal versus the paternal genes.

### Chromatin-Binding Proteins

The remaining histone methyltransferases also recognize methyl groups on other regulatory proteins; therefore, they are discussed here. The second family of SET domain proteins is related to the *Drosophila* protein Enhancer of Zeste, with the prototypical mammalian protein named EZH2. EZH2 is part of a complex of proteins called the Polycomb group (PcG). Two variants of these complexes have been designated Polycomb repression complexes 1 (PRC1) and 2 (PRC2). EZH2 belongs to the PRC2 complex that also includes EED and SUZ12; whereas PRC1 includes the proteins RNF2, HPC, EDR, and BMI1. BMI1 has received increased attention because it is an important marker of normal and cancerous hematopoietic stem cells (124–126). The Polycomb group of proteins with their SET domains not only participates in histone lysine methylation, but the complexes that they form (PRC1, PRC2) are also important in recognizing the methylated protein imprint.

A human homolog of *Drosophila* Trithorax is the mixed leukemia gene 1 (MLL1). There are four human MLL homologs. MLL1 has been shown to be a specific methyltransferase for H3 at K4 (127). In turn, it forms protein–protein interactions with coactivators, for example, CBP and corepressors chromatin remodelers (e.g., SWI/SNF) (128,129). Other Trithorax homologs (e.g., Ash1, Trx) form complexes with different coregulatory complexes. Collectively, members of the Trithorax group (TrG) of proteins can either activate or repress transcription depending on the coregulator with which they associate.

Retinoblastoma protein-interacting zinc finger protein (RIZ), SMYD3, and MDS-EVI1 form a fourth family of SET domain proteins because they have two isoforms that exhibit opposing functions. The isoform containing the SET domain has tumor suppressor function, whereas the isoform missing the SET domain is cancer promoting. This "yin-yang" theory put forth by Huang (123) is especially true for RIZ and MDS-EVI1, in which by an unclear mechanism, the cancer disturbs the normal ratio between the two isoforms. The SMYD3 protein contains another DNA-binding domain called MYND, in addition to a SET domain, and is overexpressed in colorectal and hepatocellular carcinomas (130).

Cross talk between DNA methylation and the histone modifications exists. These interactions were shown by the observation that HDAC1 forms a complex with DNMT1 and 5-methyl-cytosine binding protein (MBP) on a methylated promoter to silence gene expression (131). Similar cross talk

occurs between the HDACs SUV39 and HP1, the HDACs PRC2 and PRC1, and the HATs MLL1 and BRM (47).

### Epigenetics and Development

The epigenetic control of gene expression is a fundamental feature of mammalian development, as indicated by the occurrence of developmental arrest or abnormalities in mutants deficient in methylation or acetylation. X-chromosome inactivation is an example of sequence-identical alleles being maintained stably in different functional states. In humans, X-linked inactivation serves to normalize the level of expression of X-linked genes in female (XX) and male (XY) individuals. Mutations in genes that affect global epigenetic profiles can cause human diseases. For example, the Fragile X syndrome results when a CGG repeat in the Fragile X Mental Retardation gene 1 *(FMR1)* 5′ regulatory region expands and becomes methylated *de novo,* causing the gene to be silenced and creating a visible "fragile" site on the X chromosome under certain conditions (132). On a more global level, mutations in the *DNMT3b* gene (which regulates the DNA methylation) lead to ICF syndrome (112,114), and CBP (with acetyltransferases activity) mutations cause Rubinstei–Taybi syndrome (133).

### Epigenetics and Cancer

Epigenetic changes play an important role in tumorigenesis. The major epigenetic changes that take place during the development of cancer are generally the aberrant DNA methylation of tumor suppressor genes and histones. Chapter 17 covers in greater detail the role of epigenetic influences in cancer, but a few highlights are mentioned here to conclude this section.

Genomic methylation patterns are frequently altered in tumor cells, with global hypomethylation accompanying region-specific hypermethylation events. When hypermethylation events occur within the promoter of a tumor suppressor gene, this can silence expression of the associated gene and provide the cell with a growth advantage in a manner similar to deletions or mutations. Although cancer cells are hypomethylated in the genome compared with normal tissues, many tumor-suppressor genes are silenced in tumor cells because of hypermethylation. This aberrant methylation occurs early in tumor development and increases progressively, eventually leading to the malignant phenotype. For example, a high percentage of patients with sporadic colorectal cancers with a microsatellite instability phenotype show methylation and silencing of the gene encoding MutL protein homolog 1 (MLH1) (134). Other methylated tumor suppressors include p16CDKN2A, p14ARF, Rb, E-cadherin, and breast cancer gene-1 *(BRCA1).* Deregulation of genomic imprinting can also play a role in cancer development, as exemplified by loss of imprinting of the *IGF2* gene in Wilms' tumor (135).

Chromatin remodeling also plays an important role during tumorigenesis. Loss or misdirection of HATs has been linked to embryonic aberrations in mice (136,137) and to human cancers (138,139). Misdirection of HAT activities as a result of chromosomal translocations is associated with multiple human leukemias (140–142). In acute promyelocytic leukemia, the oncogenic fusion protein promyelocytic leukemia-retinoic acid receptor-α (PML-RARα) recruits an HDAC to repress genes essential for the differentiation of hematopoietic cells (143). Similarly, in acute myeloid leukemia (AML), AML1-eight-twenty-one (ETO) fusions recruit the repressive N-CoR-Sin3-HDAC1 complex that, in turn, inhibits normal myeloid development (144).

That many human diseases, including cancer, have an epigenetic cause has encouraged the development of a new therapeutic option called "epigenetic therapy" (145). Many agents have been discovered that alter methylation patterns on DNA or the modification of histones, and several of these agents currently are being tested in clinical trials.

## ANATOMY OF THE PROMOTER

### DNA Elements

RNA Pol II and its accessory factors bind to a DNA sequence called the *promoter,* which is located upstream of protein-coding sequences to direct RNA transcription (146). Without the promoter, the genetic sequences that encode the information to make a functional peptide product will not be transcribed. Other 5′ flanking sequences or DNA elements that participate in transcription are sequence-specific binding sites for proteins that regulate the fidelity, rate, and timing of Pol II binding, formation of the preinitiation complex (PIC), and initiation of transcript elongation under basal and regulated conditions (147–149). These sequences are defined as *cis-acting* elements because they are a part of the same (cis) gene (150–153). DNA elements are categorized according to their ability to regulate transcription as a function of their distance and orientation from the promoter. Sequences that are contained within the first 30 to 100 bp of the promoter and operate in one orientation are considered promoter-dependent, cis-acting elements. If they are *positive-acting* elements and increase the rate of transcription, they are considered *activating DNA elements,* whereas if they are *negative-acting* DNA elements and decrease or repress the rate of transcription, they are *repressor elements* (154–156).

The structure of the promoter includes several critical elements that include the TATA element, which lies upstream of the transcription start site, the initiator sequence (Inr) that spans the start site, upstream regulatory elements that bind either transcriptional activators or repressors, and finally downstream poly(dA-dT) elements (157). The TATA element, or "TATA box," is an element with a DNA sequence that is TATA or variants thereof (151,158–161). This sequence resides at a fixed distance 25 to 30 bp upstream from the transcriptional start site in many Pol II promoters, and its

location relative to the start site is dependent on position and distance (162–164). However, it became apparent that many genes did not have TATA sequences. These "TATA-less promoters" still remain dependent on assembly of the TATA-binding protein (TBP) at the promoter to form the PIC, but the recruitment of TBP is not rate limiting (165).

Inr elements, although initially identified at the "TATA-less promoters" (166,167), have subsequently been found in both TATA-containing and TATA-less promoters. Their role appears to be in directing the accuracy of Pol II initiation (168). These Inr elements reside within the first 60 bp of the transcriptional start site and directly overlap the start site itself, but they do not have a clearly defined consensus sequence (169). Many of the genes encoding GI peptides (e.g., gastrin, somatostatin, cholecystokinin [CCK], glucagons, and secretin) contain TATA elements (170–174); however, the gene encoding the growth factor, transforming growth factor-alpha (TGF-α), does not (175).

Regulatory elements are generally sequence-specific DNA elements that bind transcription factors. In the case of transcriptional activators, there are two variations, upstream activating sequences (UASs) and enhancers. Both elements are orientation and distance independent. However, UAS elements do not function downstream of the TATA box. Thus, their function is restricted by their location relative to the TATA box (176,177). UAS elements, which bind transcription factors, facilitate assembly of the PIC directly by forming protein–protein interactions with GTFs, or indirectly by complexing with coactivators. Upstream repressor sequences (URSs) use several approaches to disrupt formation of the PIC. They can interfere with the activation domain of the activator complex, disrupt interaction with the core promoter factors, or recruit corepressors (e.g., Sin3-Rpd3, HDACs). Homopolymeric dA-dT sequences are required for normal levels of transcription. The repetitive dA-dT sequence has intrinsic structural ability to impair nucleosome assembly or stability (178,179).

Models describing the formation of the Pol II initiation complex are constantly evolving and essentially involve the convergence of information gathered from biochemistry, structural biology, and genetics, particularly yeast genetics (148,180,181). Elucidation of the three-dimensional crystal structure of the TBP has advanced our understanding of preinitiation assembly complexes (182,183). Protein folding of TBP into a β-sheet forms a "saddle-shaped" concave surface of sufficient size to contact helical DNA (Fig. 1-6A). On the opposing convex surface are potential binding sites for various regulatory proteins, for example, TBP-associated factors (TAFs), GTFs, and Pol II (see Fig. 1-6A). At least 10 to 14 different human TAFs have been identified from HeLa cells, with their molecular weights ranging from 18 to 250 kDa (167,184,185). TAFs are multiple subunit proteins that associate with TBP to form the essential transcription factor TFIID. The proteins are conserved from yeast to humans with the bulk of our understanding of these factors coming from experiments in yeast and *Drosophila.* An interesting finding is that TAFs are not universally required for transcription, but each one is required for only a subset of genes. Thus, for example, one TAF is required for transcription of 8% of genes, whereas three different TAFs are required for 60% of transcribed genes. In addition, TAFs are found in protein complexes other than with TBP. In fact, some TAFs have HAT activity, whereas others are similar to histones. Still other TAFs (e.g., TAFII250) have numerous enzymatic features including ubiquitin-conjugating activity (186). The conclusion from these studies is that TAFs are involved in promoter selection through yet to be defined mechanisms (185).

TBP is not specific to Pol II promoters, but also forms PICs at the start site of Pol I and III promoters, as well as Inr promoters that do not contain TATA elements (167,187,188) (see Fig. 1-6B). For example, in Pol I promoters, TBP does not bind DNA directly in a sequence-specific manner, but instead forms protein–protein interactions with the selectivity factor complex (SL1) and the upstream binding factor (UBF) (189). In Pol III promoters, TBP complexes with TFIIIB and TFIIIC (190). In TATA-dependent and -independent Pol II promoters, TBP forms protein–protein interactions with spatially constrained upstream activators that bind DNA; for example, CCAAT-enhancer binding protein (C/EBP) and Sp1. Thus, TBP forms the core of the PIC through both DNA–protein and protein–protein contacts in TATA-dependent promoters but primarily protein–protein interactions in non-TATA promoters (see Fig. 1-6). Apparently, the selection of a promoter by TBP preceding the assembly of the PIC is determined by the type of accessory factors recruited (TAFs, SL1, Sp1, TFIIIC) (188,190,191). Moreover, this recruitment may be regulated by temporal and tissue-specific influences. Inhibition of transcription (repression) may occur simply by preventing one of these general TAFs from participating in the assembly of the PIC (182,187). Like Pol II itself, many TAFs and GTFs are composed of multiple subunits. Thus, there is an enormously complex pattern of assembly of proteins (TBP + TAFs = TFIID, other GTFs, and upstream activators) on specific DNA elements at the promoter (e.g., TATA, INR, UAS) that results in the initiation and elongation of mRNA (182,188,192).

Other GTFs besides the TFIID complex include TFIIA, TFIIB, TFIIE, TFIIF, TFIIG, TFIIH, TFIII, TFIIJ, and TFIIK (160,167,193–195). There appears to be a strict requirement for these factors to assemble at the promoter in a specific order (181,182,192). TFIID binds to the TATA elements first, followed by protein–protein interactions of TFIID with TFIIA and TFIIB. The 12-subunit Pol II binds next. TFIIF is then recruited to the TFII-diaminobenzidine complex and facilitates binding of other general (basal) transcription factors E, J, H, and K. Many of these basal factors do not bind DNA directly (e.g., TFIIB, TFIIE, TFIIF), but instead form bridging complexes between the general Pol II transcriptional machinery and TAFs with specific upstream regulators. GTFs are required for the basal activity of the promoter, whereas UAS enhancers are dispensable.

Specific functions of some of the GTFs have been elucidated. For example, the larger subunit of TFIIF (Rap74)

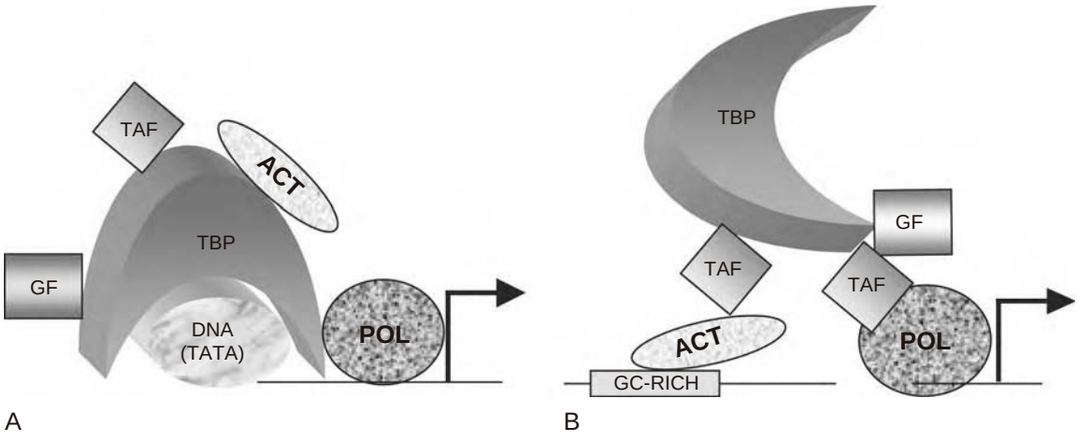



**FIG. 1-6.** Schematic diagram of a polymerase (POL) II initiation complex. The saddle-shaped TATA-binding protein (TBP) **(A)** binds DNA directly at the TATA sequence and **(B)** is tethered between TBP-associated factors (TAFs) in non-TATA promoters. Thus, TBP forms the core of a complex consisting of TAFs, general transcription factors (GF), upstream activators (ACT), and ribonucleic acid (RNA) POL I, II, and III. (Modified from Comai and colleagues [187], by permission.)

functions as an ATPase-dependent helicase to unwind DNA ahead of the transcription complex (196). TFIIF appears to play a role in promoter stability rather than selectivity. TFIII, a helix-loop-helix (HLH) protein, binds preferentially to Inr promoter with or without TATA elements and cooperates with upstream regulatory factors and the general transcription complex (197). TFIIH is one of several C-terminal domain (CTD) kinases that phosphorylates the CTD of Pol II to signal elongation of the nascent mRNA chain (192,194). Other kinases are now known to phosphorylate CTD (198).

Certainly, all genes are not transcribed concurrently; thus, the cell must have various mechanisms for silencing genes either permanently or in response to extracellular cues. The mechanisms for repressing genes may be general (e.g., DNA methylation [199,200]; see also #1375 in Bird [201]) or sequence-specific (202). Alternatively, loss of the ability to inhibit transcription of a gene (derepression) may permit certain cellular functions to proceed unchecked. Examples of the interaction between positive and negative regulators occur during cellular proliferation and differentiation (203). During fetal development, most cells are in the process of rapid proliferation. This period is followed by one of regulated differentiation during which the genes controlling proliferation are repressed. However, proliferative pathways may be derepressed (reactivated) during periods of organ repair or during neoplastic transformation (203,204). Examples include the reexpression of fetal proteins during liver regeneration and neoplasia (e.g., α-fetoprotein) or GI mucosal neoplasia (e.g., carcinoembryonic antigen) (205–207). Negative promoter elements or repressors may serve as the binding sites for proteins that sterically hinder the binding of GTFs (e.g., TFIID) or upstream activators (e.g., Sp1) critical in the formation of the Pol II transcription PICs (DNA–protein interactions). Alternatively, proteins responsible for gene repression may act by preventing the recruitment of required general or accessory factors (e.g., TFIIB or TAFs) to the bound PIC (protein–protein interactions) (202).

The DNA elements CCAAT and GGGCGG, which bind the nuclear proteins C/EBP and Sp1, respectively, are examples of promoter-activating elements that are distinct from the TATA box (151). These *upstream promoter elements* are distinguished from the TATA element in that mutation or removal of these UASs reduces basal promoter activity without completely eliminating it, whereas mutation or elimination of the TATA sequence completely abolishes transcription. DNA elements that function independently of their position on the gene or their orientation (3′ to 5′ or 5′ to 3′) are called *enhancers* if they bind nuclear proteins that activate transcription and *silencers* if they bind nuclear proteins that inhibit transcription (208–210). Many of these enhancer and silencer elements occur far upstream within the 5′ flank, but they may also occur within introns, exons, or 5′ or 3′ untranslated sequences.

To identify cis-acting enhancer elements, constructs are made by ligating the regulatory elements to be studied in front of a functional promoter expressing a gene encoding a protein or enzyme that is easily assayed. Typical reporter genes encode proteins that are not normally expressed by the transfected cell. By systematically deleting portions of 5′ flanking sequence, the transcriptional activity of the promoter under various conditions is altered and the regulatory elements of interest are identified. DNA elements responsible for tissue specificity can be identified by transfecting (transferring DNA into eukaryotic cell lines) cell lines derived from different tissues. Transcriptional initiation from a promoter that requires a particular cis-acting sequence for expression in a specific cell type is diminished or abolished if this sequence is eliminated or mutated.

Cis-acting sequences conferring inducible responses are also identified by this method. Alternatively, elements that are only active during development must be identified in eukaryotic systems in which differentiation of a cell line can be controlled, or in transgenic animal models.

### DNA-Binding Proteins

DNA-binding proteins are also referred to as *trans-acting factors* by virtue of their ability to bind to the 5′ flanking regions of genes in a sequence-specific manner and regulate transcription (211–213). The term *trans* was coined to acknowledge that the protein product of one gene regulates the transcription of a different gene. With the genes for several hundred trans-acting factors now cloned, the study of their primary and secondary amino acid structures has demonstrated characteristic protein domains (214,215). In general, these proteins contain specific DNA-binding, transactivation, and oligomerization domains (Fig. 1-7). The amount of a transcription factor binding to a particular sequence initially is considered to be the primary mechanism of control. However, it is now clear that the proteins themselves are regulated by a variety of mechanisms in addition to controlling their levels in the nucleus and include activation or inactivation by proteolysis (e.g., NF-κB), covalent modification (e.g., phosphorylation, acetylation), and ligand binding (e.g., steroid receptors), in addition to regulating translocation to and from the nucleus and transcriptional induction or repression of the trans-acting factor (216).

The DNA-binding domain is the portion of the protein that contacts DNA in a sequence-specific manner. However, flanking amino acids may also influence DNA-binding through noncovalent interactions. Examples of four major designs for DNA-binding domains are proteins with a helix-turn-helix domain, "zinc finger" domains, amphipathic helices (e.g., basic-zipper [bZip], HLH), and β-ribbon (prokaryotic proteins) (215) (Fig. 1-8). Most of the protein-DNA contacts occur in the major groove through noncovalent interactions (e.g., hydrogen bonds, hydrophobic interactions, and van der Waals interactions). An α-helical structure appears to be a common motif used in the formation of the DNA-binding domain.

The helix-turn-helix motif was initially identified in prokaryotic DNA-binding proteins, but similar motifs have now been identified in the homeodomains of eukaryotic transcription factors (217–219) (see Fig. 1-8). Homeobox factors are a class of DNA-binding proteins that predominantly play a role in the developmental expression of genes. Their discovery arose from the idea that developmental regulation involves control of gene expression by a few regulatory transcription factors called "master switch genes" (220). These DNA regulatory proteins initially were identified in simpler organisms such as the roundworm *Caenorhabditis elegans (C. elegans)* or *Drosophila,* in which the genetic development from the single-cell stage to maturity is well defined.

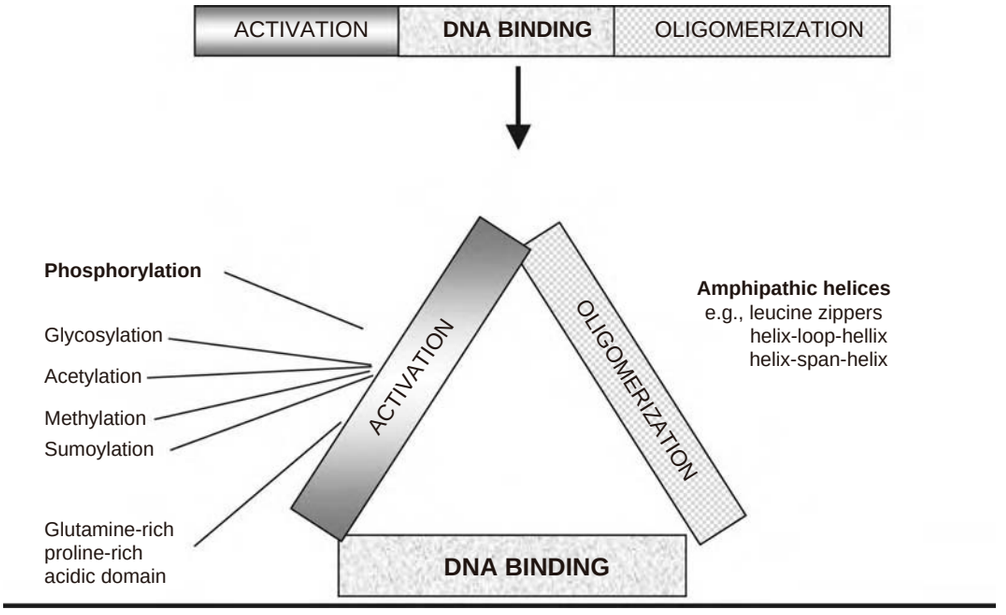



**FIG. 1-7.** DNA-binding protein domains. DNA-binding proteins have three functional domains: (1) a surface to bind DNA, (2) a surface to interact with other proteins containing similar oligomerization (dimerization) motifs, and (3) a surface to interact with the signal transduction pathways or the preinitiation complex. Examples of motifs associated with these domains are listed.

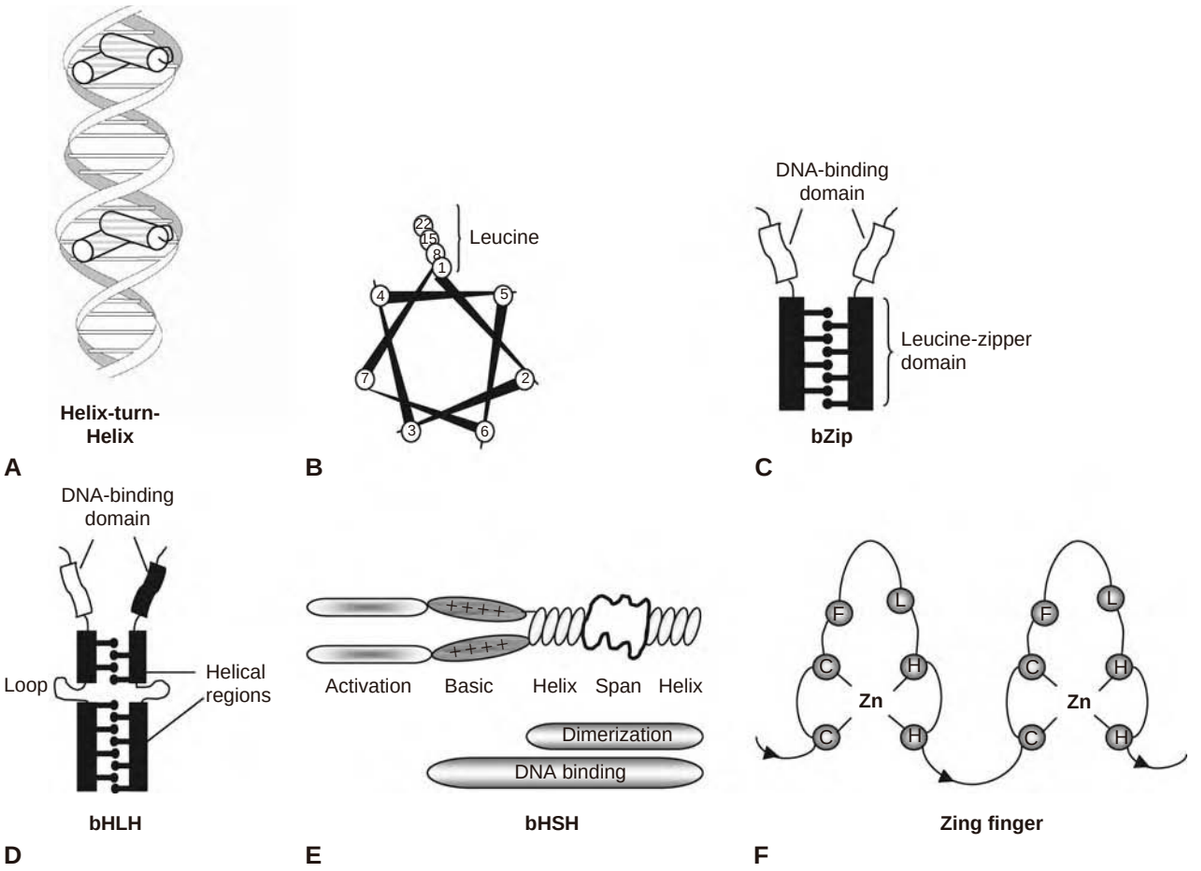



**FIG. 1-8.** Example of DNA-binding motifs. **(A)** Helix-turn-helix. **(B)** Amphipathic helix formation. An α-helical structure is formed so that all leucines or hydrophobic amino acids line up at one surface. Shown are **(C)** a leucine zipper adjacent to basic DNA-binding domain (bZip); **(D)** a helix-loop-helix adjacent to a basic DNA-binding domain (bHLH); **(E)** and a helix-span-helix adjacent to a basic domain (bHSH). **(F)** A zinc finger motif of the C2H2 type with tetrahedral coordination of a zinc ion between two histidines and two cysteines. (Modified from Berg [241], Falke and Juliano [255], Ellenberger and colleagues [266].)

Through site-directed mutagenesis studies, a specific protein domain required to effect developmental progression of these organisms was identified. This domain shared significant homology with a region within proteins controlling cell lineage in the pituitary (Pit-1) and immune system (B-cell octamer proteins, Oct-1 and -2) (221–225). These proteins also shared significant homology with the *C. elegans* "homeotic gene," *unc-86*. Thus, the ~60- to 75-amino-acid region of shared homology was renamed the "POU" domain after the three proteins Pit-1, Oct-1, and unc-86. Initially, the POU domain was named without knowledge of function or the ability to form specific secondary protein structure. Although some bind similar AT-rich consensus DNA-binding sites (octamer proteins bind an eight-nucleotide sequence ATTTGCAT; Pit-1 [also called GHF-1] binds a nine-nucleotide consensus site $^{T}_{A}{}^{T}_{A}$TATNCAT), others do not (*Drosophila eve* protein recognizes TCAGCACCG) (217). Mouse homeobox genes have nomenclature based on their similarity to *Drosophila* homeobox genes (e.g., *caudal*, *forkhead*) and have been associated with control of gut development (226–228).

In fact, homeobox genes have emerged as critical regulatory factors in the development of both the luminal GI tract and pancreas (229,230). Homeobox genes in the luminal GI tract are related to the 39-member *Hox* gene family of transcriptional regulators that control anterior-posterior patterning, and they are related structurally to the *Drosophila Antennapaedia* gene (229). *Hox* genes are so strongly conserved in evolution that this cluster of genes has been repeated four times in mammals on different chromosomes (231). Collectively, the replicated genes are called Hox clusters and are expressed primarily in either the mesoderm or ectoderm (e.g., skin, muscle, neural tissue), but not in endodermal tissues. Rather, an evolutionarily related cluster of homeotic genes call the *Para-Hox* genes appear to play the more important role in endodermal tissue, and therefore gut patterning (232). These genes include *Pdx1,* which is

essential to the correct development of the pancreas and duodenum (233,234), and the genes related to the *Drosophila caudal* gene, *Cdx1, Cdx2,* and *Cdx4* (229,235). *Cdx2* is not only relevant to development of the luminal gut, but it also is an indicator of neoplastic transformation, especially in the upper GI tract (236,237). The *forkhead* family of homeotic genes is another group of transcriptional regulators with important implications in the gut because of their role in GI cancers (238,239). There are at least 43 members of the *forkhead* family spread over three chromosomes. The "winged helix" motif of the forkhead DNA-binding proteins is a variant of the 60-amino-acid homeodomain helix-turn-helix because it has additional peptide domains that have been described as "wings" (240). The forkhead transcription factors are downstream targets of the hedgehog pathway, which is an important developmental signaling cascade originally described in *Drosophila* (see Chapter 9 for a more detailed discussion).

The zinc finger motif is distinguished by the occurrence of cysteine and histidine resides tetrahedrally coordinating a zinc ion (241–243) (see Fig. 1-8F). Two subcategories of zinc finger proteins have been identified: those regulatory proteins in which only cysteine contacts the zinc ion (e.g., the steroid receptor family, GAL4 [244,245]), and those in which both cysteine and histidine residues are involved (e.g., Sp1 and Zif 268 [245,246]). The X-ray crystallographic structures of several zinc finger and helix-turn-helix proteins have now been identified, with identification of more structures still to follow (247–253). Through crystallographic studies and computer modeling, investigators have been able to identify which amino acids within the DNA-binding domain contact particular nucleotides within the DNA element. It is anticipated that most of these interactions will be defined sufficiently well to predict protein–DNA contacts at the molecular level for other trans-acting factors. In the future, this will facilitate the targeting of specific transcription factors (natural or synthetic) to specific promoter sequences (248,254–257).

Landschulz and coworkers (258) originally described the bZip/coiled-coiled DNA-binding motif as a dimerization domain (see Fig. 1-8). However, this motif, which consists of 55 to 65 amino acid residues, actually forms two domains: one for dimerization and a second for DNA binding (259). Seven repeating leucine residues forming an α-helical coil compose the dimerization domain (Zip domain) (see Fig. 1-8). Immediately adjacent to the Zip domain, toward the amino terminus, lies the basic/hydrophobic domain (b domain) (215). Thus, the bZip family of proteins, the first of the amphipathic helices to be described, must dimerize to form a complete DNA-binding domain (260,261). Other transcriptional regulatory proteins containing the same heptad repeat are able to dimerize with each other to form a "coiled coil" (262). For stable binding to DNA to occur, some bZip proteins prefer that each dimerization partner be the same (e.g., CREB, C/EBP, or general control of amino acid synthesis 4 [GCN4] homodimers [see #564 in Pu (260); 263,264]), whereas other bZip proteins form more stable complexes as heterodimers (e.g., Fos/Jun), although lower affinity binding is also possible as homodimers (e.g., Jun/Jun) (265). The first report of a crystal structure for a bZip protein, the yeast transcription factor GCN4, confirmed the predicted model of two α-helical coils, which merge into diverging b domains that straddle and grip the major groove of DNA like "forceps" (266).

Other amphipathic helices, which combine a dimerization domain with a basic DNA-binding domain, have been described; however, less is known about their three-dimensional structure. The helical domains contain hydrophobic amino acid residues arrayed in an α helix so that they are clustered on one face of the helix, whereas hydrophilic residues reside on the opposing face (see Fig. 1-8). According to thermodynamic principles, the hydrophobic face is sequestered away from the aqueous environment by noncovalent interactions when they dimerize with similar domains on other proteins. In addition to the bZip model described earlier, the HLH and helix-span-helix (HSH) motifs were coined to describe other subclasses of amphipathic helices, albeit with longer linker sequences between the two α helices (267–269) (see Fig. 1-8). In the case of the leucine zipper, the hydrophobic face is formed by a series of leucine residues spaced seven amino acids apart (258). In contrast, the HLH and HSH proteins use a variety of different hydrophobic amino acids in addition to leucine to form two amphipathic α helices separated by a stretch of amino acids ("loop or span") that do not form a helix. Like the bZip family, HLH and HSH regulatory proteins bind DNA through an adjacent basic domain.

Thus, bZip proteins (e.g., CREB, activator protein 1 [AP1], activating transcription factor [ATF], Fos, Jun) are potentially interchangeable partners within homodimeric or heterodimeric complexes with the corresponding ability to recognize a greater repertoire of DNA-binding elements (270–272). For example, the Fos/Jun-binding site differs from the CREB/ATF-binding site by 1 bp: CREB/ATF binds TGACGTCA, whereas Fos/Jun binds TGAGTCA. Likewise, the bHLH proteins that recognize the CANNTG consensus binding site are also able to complex with each other (273). Currently, there are three family members of the transcription factor AP2, which are the only members of the bHSH family (269, 274,275). An HLH protein without the basic DNA-binding domain called Id was cloned (276). This protein has been shown to combine with three bHLH proteins (MyoD, E12, and E47) and to prevent the formation of normal homodimers or heterodimers, thereby functioning as a dominant negative mutant. Similar types of negative regulatory proteins have been identified for bZip proteins (277,278). Therefore, the combinatorial ability of transcription factors permits flexibility in responding to extracellular signals at the level of DNA–protein and protein–protein interactions.

The transactivation domains of regulatory proteins consist of predominantly acidic, basic (glutamine), or proline residues (152,279). These non-DNA-binding surfaces interface with signal transduction pathways and other proteins, but their specific function is not completely understood. Domains with a high degree of acidic charges are thought to represent important contact points for interaction with the Pol II PIC

(e.g., Gal 4, VP16). Ptashne (280) coined the phrase *acidic blobs* to describe such negatively charged trans-activating domains. Glutamine-rich (Sp1) and proline-rich (C/EBP, CTF) domains also presumably cooperate with the transcriptional machinery through protein–protein interactions (160,184,281–284). However, it has more recently been confirmed that transcription factors form protein–protein interactions with other transcription factors not within the same DNA-binding domain family. The most common transcription factor exhibiting this property is Sp1. Sp1 can interact directly with other transcription factors, for example, YY1, Smads, or Jun family members (285,286). A functional interaction between cJun and Sp1 has been shown to mediate epidermal growth factor activation of lipoxygenase gene expression (287). Presumably, the "acidic blob" in the transactivation domain of Sp1 creates a "sticky" surface on which new partnerships are formed at various promoters in response to a variety of extracellular signals. Likewise, Smad proteins, which mediate TGF-β signaling, are also promiscuous in their ability to partner with other transcription factor family members (288,289). Although at one time undetected, protein–protein interactions among transcription factors are now recognized as common occurrences, particularly because there are convenient means to identify the interactions genetically through two-hybrid cloning methods, or biochemically using affinity chromatography, immunoblot assays, and mass spectroscopy.

Many of the mechanisms involving transactivation of transcription factors involve protein phosphorylation and dephosphorylation (290). Phosphorylation by protein kinases occurs at serine, threonine, or tyrosine amino acid residues. Several classes of protein kinases exist within the cell; however, the best studied are the protein kinase A (PKA) and C (PKC) pathways. PKA is activated indirectly by the catalytic subunit of adenylate cyclase. Signals that increase intracellular cAMP will activate PKA (291,292). In contrast, PKC is activated by calcium released from intracellular stores and by the phospholipid diacylglycerol (293). Phospholipase Cγ catalyzes the hydrolysis of phosphatidylinositol to diacylglycerol. The tumor promoter 12-0-tetra-decanoyl phorbol-13-acetate (TPA) is a lipid-soluble compound that mimics diacylglycerol and directly activates PKC. Hundreds of additional protein kinases within both the cytoplasm and the nucleus exist that may be implicated in the specific phosphorylation of transcription factors (294). Ligand binding triggers a variety of different activation pathways that appear to result in the direct phosphorylation of transcription factors by protein kinases other than PKC and PKA; for example, casein kinase II (CKII), glycogen synthase kinase III, and several DNA-dependent protein kinases (295–297). Direct phosphorylation of the DNA-binding protein may result in a conformational change that enhances its ability to induce transcriptional activation (e.g., CREB, cJun, C/EBP-β) or inhibition (e.g., yeast protein A[298]DRI) (290). Alternatively, phosphorylation of an inhibitory subunit may release the transcription factor from an inactive state (e.g., NF–κB) (299–303). Phosphorylation can also regulate the ability of a protein to dimerize, thereby broadening or narrowing the repertoire of DNA sequences that are recognized (e.g., signal transducer and activator of transcription [STAT] and Fos/Jun family) (304,305).

The removal of phosphate groups by sequence-specific phosphatases is an additional mechanism by which the transcriptional activity of DNA-binding proteins may be altered (290,306,307). Interestingly, dephosphorylation appears to be a more common mechanism for regulating transacting factor binding than is kinase-mediated phosphorylation (295). Binding of the Jun family (bZip class), homeodomain proteins, and cMyb to DNA is regulated by dephosphorylation. Phosphorylation of sites within or adjacent to the DNA-binding domain of these proteins inhibits DNA binding, whereas removal of phosphates enhances binding. In contrast, activation of DNA binding by phosphorylation has fewer documented examples. One example is the serum-response factor (SRF) that binds to and activates the cFos promoter (308,309). SRF appears to be activated by phosphorylation at sites adjacent to the DNA-binding domain by CKII. This observation is supported by studies involving both mutational analysis of these phosphorylation sites and increasing cellular CKII kinase activity through microinjection of the enzyme into cells (310,311).

Although glycosylated proteins are usually observed on the plasma membrane of cells or in the lumen of intracellular organelles, nuclear proteins have been shown to contain O-linked glycosylated residues as well (312). Sp1 represents the prototypical glycosylated transcription factor, the activity of which is enhanced by the presence of carbohydrate residues (312–315). Other eukaryotic transcription factors such as CTF, AP1, and AP4 are also known to be glycosylated, but the effect of the carbohydrate residues on their transcriptional activity is unknown. Glycosylation may regulate the transcriptional activity of individual transcription factors, perhaps by increasing their resistance to proteolysis, by targeting them to the nucleus, by blocking potential phosphorylation sites, or by facilitating their interaction with coactivators (316).

### Coregulatory Proteins

By the mid 1990s, it became clear that DNA-binding factors were working in a combinatorial manner, not only with other DNA-binding factors, but with non-DNA-binding proteins that were closely linked to chromatin structure and the PIC. These large molecular weight proteins were initially identified as factors interacting with the steroid hormone receptors, which are DNA-binding proteins that translocate to the nucleus after binding hydrophobic ligands in the cytoplasm (317–319). At about the same time, it was discovered that phosphorylation of the cAMP-activated transcription factor CREB induced its interaction with a 300-kDa coactivator protein called CBP. Subsequent to the discovery of CREB, the homologous transcriptional coactivator designated p300 was also identified (320). Coactivators were

found to facilitate transcriptional activation through intrinsic HAT activity, resulting in an "open" chromatin state at the start site of transcription. There are now several of this class of proteins that include PCAF and GCN5 (321). Conversely, the protein complexes that inhibited transcription were multiprotein complexes that recruited histone deacetylators, which, in turn, deacetylate histones returning chromatin to its closed, inactive state (322). The prototype corepressors were identified because of their ability to suppress activation by the retinoid and thyroid hormones (SMRT/N-CoR) (216). It is now known that there are transcriptional corepressors of a variety of signal transduction pathways, including Sin3A, a corepressor of the cMyc bHLH transcription factor family, and PIAS/SUMO, a corepressor of the STAT signaling pathway (323–328).

Collectively, these proteins are considered to be coregulatory factors because they do not contact DNA directly as transcription factors do, but rather form protein bridges between the sequence-specific DNA-binding proteins and the Pol II assembly apparatus, bringing with them enzymatic activity, for example, acetylase and deacetylase activity involved in remodeling chromatin (329–331). Currently, there are three broad categories of coactivators (332). p300 and CBP are the prototypes of the HAT class of coactivators. The TRAP/DRIP/Mediator/ARC complex compose the second class and are proteins that bind transcription factors and recruit RNA Pol II without having intrinsic histone modification capabilities. The third class comprises the yeast SWI/SNF and their mammalian homologues BRG1/BRM. This third class of coactivators contains intrinsic ATP-dependent DNA-unwinding activity required for efficient *in vivo* transcription. Coactivators increase the transcriptional activation of a promoter through its interaction with a sequence-specific DNA-binding protein, but it is not yet clear how the coactivator selects one group of promoters over another. Two concepts have been considered (332). For example, a promoter might need a "threshold level" of positive signals to be activated. Alternatively, some promoters might have a greater requirement for the presence of one coactivator than another.

The precise mechanisms of transcriptional activation continue to evolve, and certain themes are emerging. In rare instances, positive or negative enhancer activity is dependent on a single DNA-binding protein that functions as a master switch to activate a family of related genes, for example, the myogenin MyoD family in muscle differentiation (333). However, further scrutiny of this model has indicated a large network of transcription factors that interact with non-DNA-binding complexes involved in chromatin remodeling, for example, histone acetyltransferase proteins p300 and CBP (334–337). Therefore, the more common mechanism implies that most cells respond to their environment by recruiting subsets of ubiquitous and promoter-specific transcription factors that combinatorially produce the desired cellular phenotype (204,338-341). Corepressors SMRT and N-CoR both recruit HDACs, yet they mediate activation downstream of different kinase cascades (342). In addition to the recruitment of classic HDAC-associated corepressors (e.g., mSin3A and Groucho) (343–345), the runt-related transcription factor (RUNX) proteins exert gene silencing by associating with histone methyltransferases (e.g., SUV39H1) (346). Bifunctional attributes of transcription factors have been attributed to their regulated association with either coactivators or corepressors.

## METHODOLOGY

This section summarizes some of the molecular techniques used to study transcriptional control of genes. These methods are used to study either genetic structure or function. Three systems have been used to study function: reconstituted cell-free transcription assays, cell culture models, and whole-animal studies. Methods that analyze structural interactions include those techniques that assess DNA–protein interactions and those that assess protein–protein interactions.

### Functional Methods

#### *Reconstituted Transcription Systems*

The most basic approach to the functional study of a gene is an *in vitro* transcription system in which the minimal components required for transcription are isolated and reconstituted to produce the gene product (347,348). mRNA is transcribed from cloned cDNA in the presence of radiolabeled nucleotides, RNA polymerase, and accessory factors isolated from nuclear extracts. The radiolabeled RNA synthesized *in vitro* is resolved by gel electrophoresis after extraction from the cell. Changes in basal levels of transcription are measured by quantifying the amount of newly synthesized RNA transcripts produced in the presence or absence of cloned or purified gene-specific DNA-binding proteins (349). In this way, differences in gene expression attributable to the activity of a purified transcription factor or enriched nuclear fraction may then be studied under tightly controlled assay conditions.

#### *Cell Culture Models*

The study of transcriptional regulation has been advanced greatly by the use of cell lines derived from the same tissues as the endogenous gene of interest. These cell lines have become the vehicles in which the study of gene expression is performed. Two major advantages of using cell lines are that they are homogeneous populations and they continue to divide in minimal culture conditions. However, in many situations, the cell lines are derived from neoplastic tissues, which may have lost the normal regulatory mechanisms that maintain the differentiated state. In a dedifferentiated state, cells tend to express a variety of genes outside of the repertoire expressed by their normal counterparts. Therefore, studies with cell lines always carry the caveat that they may not reflect activities of native cells.

The use of cell lines permits the direct study of regulators of endogenous gene expression, avoiding the confounding effects of contaminating cell types. However, this approach does not permit alteration of the regulatory domains of genes to assess their contribution to transcription. Therefore, techniques have been developed to insert altered genetic material into cells by chemical, electrical, or viral mechanisms. In this way, specific elements controlling transcription can be isolated and studied. To tag the inserted gene, the promoter from which transcription will be initiated is ligated upstream of the coding sequences for a reporter gene, for example, chloramphenicol acetyltransferase, β-galactosidase, growth hormone, green fluorescent protein, or luciferase (9,350–353). The products of the reporter gene are easily measured, and spurious detection of reporter gene activity is kept to a minimum because their products are not normally expressed by most mammalian cells. Regulatory sequences to be analyzed are ligated upstream of a promoter with basal transcriptional activity in the test cell line. Taking advantage of various restriction sites, sequentially shorter 5′ flanking sequences are created, and each resulting construct is then tested by assaying the reporter gene product as an indicator of gene expression.

### *Whole-Animal Models*

Whole-animal studies have been useful in assessing the contribution of transcriptional control to the regulation of several GI peptides, including gastrin, CCK, and somatostatin (354–356). Brand and Stone (357) showed that gastrin mRNA levels in the antrum increase under conditions of chemical or surgical achlorhydria and coincide with a reciprocal decrease in somatostatin mRNA. These observations are correlated with prior observations that gastrin plasma levels increase under conditions of achlorhydria (354, 358). Furthermore, infusion of the somatostatin analogue octreotide blocks the increase in gastrin mRNA (357). Walsh and coworkers (359,360) found that gastrin mRNA levels are predictably regulated by cycles of fasting and refeeding. Recently, infusion of the proinflammatory cytokine interferon-γ into mice has been used to recapitulate the effect of *Helicobacter pylori* infection on gastrin and somatostatin (361). Similarly, studies on the dietary control of CCK gene expression have been reported (355). Although such studies permit the linkage of transcriptional regulation to physiologic events, they do not allow dissection of the responsible regulatory elements.

### *Transgenic Animals*

Through transgenic animals it is possible to introduce genetic information into the mouse genome such that there is permanent alteration of the genetic makeup in both the founder line and successive generations (13,14). Transgenic studies afford the opportunity to study the importance of specific genetic sequences in cell, organ, and whole-animal function. By breeding mice with different transgenic lineages, the interaction between these artificially produced genotypes on the overall phenotype may be amplified or abolished. In many situations, these alterations reproduce clinically relevant pathologic states (362–366). Chapter 53 provides specific details on transgenic technology including the powerful technique of homologous recombination.

### *Cell-Based Knockout Strategies*

Once a genetic target is identified, whether DNA, RNA, or protein, the next step is to determine the significance of the molecule in a particular signaling, developmental, or neoplastic cascade. This usually is done by blocking, reducing, or removing the gene product at the cellular level before applying the extracellular signal. A change in the expected phenotype would confirm that the gene product makes a significant contribution. At the cellular level, the traditional approach has been to use small molecules, for example, pharmaceutical inhibitors. Once DNA vectors were developed in the early 1980s, antisense and dominant negative approaches to inhibit gene expression came into vogue (367). With the emergence of transgenic technology, it became apparent that one could remove the gene product through genetic manipulation specifically by homologous recombination to disrupt the gene in mice (362,368). With the discovery of snRNA molecules that interfere with either transcriptional initiation or translation, the commercial availability of synthetic "interfering" RNA molecules has emerged (369,370). High-throughput methods using RNA silencing are now being used to complement the gene discovery methods of DNA microarray technologies (371). Nevertheless, RNA interference technology, although relatively easy to use, does not eliminate the gene product as effectively as direct gene targeting. Therefore, genetic methods must be used to generate a complete null cell line. Cell lines are either created from a null mouse model (e.g., embryonic fibroblasts), or somatic cell gene targeting can be performed in the cell line of choice (372,373). The advantage of creating the null cell line from a mouse is that the cells will be from normal tissue and not a tumor cell line. However, unless molecules are introduced to immortalize the cells, the lines are not permanent. Gene targeting in a somatic cell line has not been as widely used because of the difficulty in performing the technique, but it is a powerful approach that permits the study of a null locus without incurring the expense of mice.

## Structural Methods

Once functional regulatory DNA elements have been identified, assays that assess DNA–protein interactions are performed (374). Indeed, in circumstances where a long sequence (>50 bp) must be analyzed, it is simpler to identify DNA–protein interactions first, and then determine whether these DNA elements are involved in transcriptional regulation. DNase I footprinting assays are used to identify DNA-binding elements that interact with crude or purified nuclear proteins by protecting them from chemical or enzymatic

cleavage (375,376). Such assays are particularly well suited for studying cooperative interactions among proteins bound to adjacent DNA elements. The technique can be performed *in vivo* or *in vitro* (9). However, *in vivo* footprinting has been superseded by chromatin immunoprecipitation (ChIP) assays (see the next section). Electrophoretic gel mobility shift assays (EMSAs; gel shift, gel delay, or band-shift assays) permit a more detailed analysis of the following: (1) the type of protein complexes that bind to individual DNA elements, and (2) the specificity of the protein interaction with a specific base pair (377–379) (Fig. 1-9). This assay system is also rapid and easier to use than footprinting assays. Methylation interference assays extend the power of the gel shift assay by identifying specific nucleotide contacts that are required for DNA binding (380). DNA affinity precipitation is a DNA–protein interaction assay that uses the biotinylated DNA binding site to identify the proteins that are recruited to the element (381). The assay uses the DNA element to isolate the protein factors, coupled with immunoblots to identify the proteins that form both the protein–DNA and protein–protein interactions. Southwestern blot analysis takes advantage of specific DNA elements that are used to detect nuclear proteins separated on a denaturing gel and transferred to nitrocellulose or produced by a phage expression library (382–384).

### *Chromatin Immunoprecipitation Assays*

ChIP analysis is now the most effective method to document an *in vivo* interaction at DNA (385–387). First, a fixative, usually formaldehyde, is used to cross-link proteins to DNA. Then antibodies are used to immunoprecipitate the DNA-binding proteins. After a series of extractions to remove the protein from DNA, specific primers are used to PCR amplify the DNA-binding element precipitated with the protein and antibody. Variations of this method are used to identify the *in vivo* preferred binding sites of known DNA-binding proteins. Alternatively, the immunoprecipitate is resolved on a sodium dodecyl sulfate gel, and mass spectroscopy can be used to identify the proteins that coprecipitate and are likely involved in protein–protein interactions with the DNA-binding proteins. The technique completely depends on the quality of the antibodies, the quantity and quality of genomic DNA precipitated, and primer specificity. ChIP assays complement *in vitro* DNA–protein interaction assays such as EMSAs or footprinting. Expression vectors or cell-based knockout strategies using dominant negative constructs, antisense technology, or RNA interference may be used to demonstrate functional significance (388). These approaches are rapid and useful to perform before using transgenic mouse approaches.

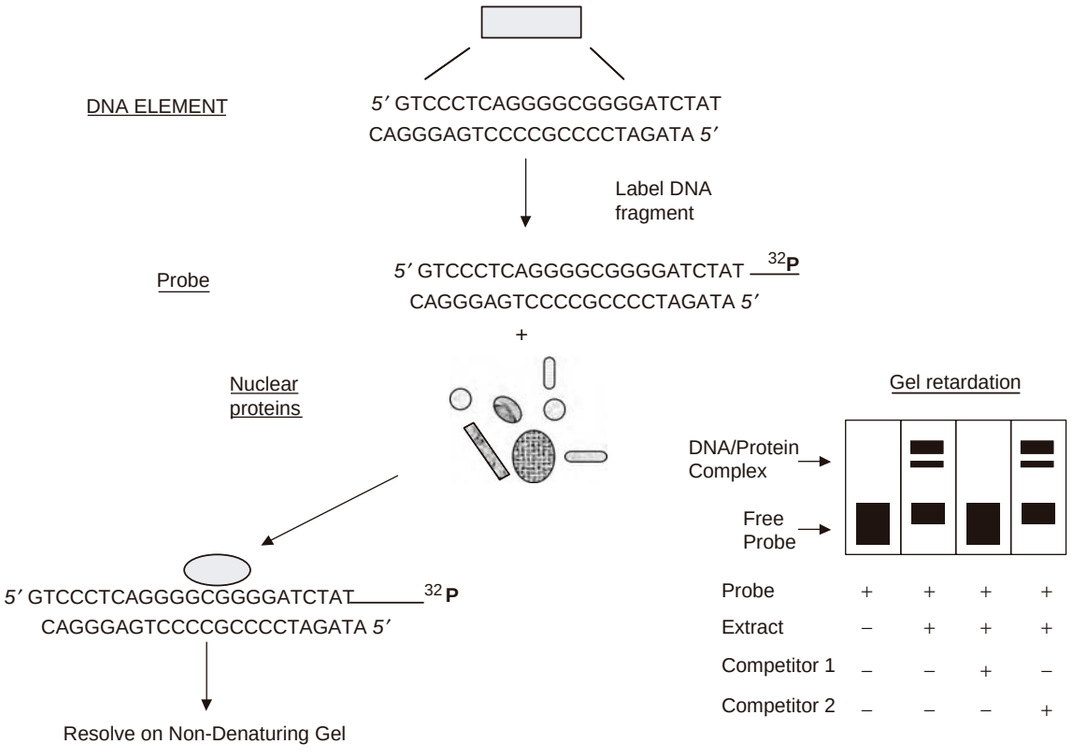



**FIG. 1-9.** Electrophoretic mobility shift assay (EMSA, gel shift). A DNA element ~30 to 100 bp in length is labeled, and then is incubated with crude nuclear extract or purified protein. A band on the autoradiogram is detected if the radiolabeled probe is retarded and does not migrate to the bottom of the gel. The specificity of binding is determined by competing with unlabeled DNA sequences. Competitor 1 is related to the probe sequence, whereas Competitor 2 is unrelated to the probe sequence.

### *Microarray Technology*

The latest method to comprehensively analyze gene expression is by microarray technology. At the transcription level, DNA array technology increases by several orders of magnitude the number of genes that can be examined simultaneously under different conditions (389–391). The number of genes that are either stimulated or inhibited under various conditions can be studied simultaneously with the limitations being the number of genomic sequences that are spotted on the glass slide. A glass slide is able to hold the genomic sequences of 25,000 to 30,000 genes, which is the current estimate of the total number of genes in the human genome. Two types of arrays are available: EST/cDNA and oligonucleotide (Affymetrix, [Santa Clara, CA]) based. The EST microarray chips use *e*xpressed *s*equence *t*ags that are fragments of DNA corresponding to segments of the genome that encode mRNA. The Affymetrix gene chips spot commercially designed oligonucleotide sequences. These DNA fragments are subsequently "arrayed" onto glass slides. In most instances, several regions of the genomic sequence unique to that gene are spotted in multiple copies to ensure reproducibility. Different genetic domains are plated because of differences in hybridization affinity. RNA is isolated from cells or tissue after treatment with an extracellular molecule or from cells at different stages of development or transformation. cDNA are then generated and tagged fluorescently, then hybridized under stringent conditions to the DNA arrayed on the glass slide followed by analysis by a special plate reader. Computer-generated algorithms are required to interpret the fluorescent signals and rank the degree of change from baseline fluorescence. The technology is being used to study the gene expression pattern found in various tissues at designated stages, for example, developmental or transformation stages (392–394). The significance of the findings must be confirmed by alternative methods including Northern blot analysis or quantitative PCR.

### *Proteomics*

Analogous high-throughput approaches have been developed to study protein modifications (395). However, the techniques used to detect protein posttranslational modifications are more complex and use more labor-intensive technology. Protein is extracted from the cell or organelle of interest and resolved by two-dimensional gel electrophoresis, in which proteins are separated by both size and ionic charge (along a pH gradient). The proteins are visualized with a dye either directly on the gel or after transfer to a paper substrate. Both substrates (gel or paper) can be used for further analysis. However, proteins transferred to a paper substrate permit several options for analysis. Resolved proteins that are transferred to paper can be submitted for analysis with an antibody (immunoblot) that might recognize phosphorylated or acetylated peptides. Differences in the size of the spot corresponding to the amount of a particular protein version (phosphorylated, acetylated) can be quantified by computer. Proteins that cannot be identified by antibody can be analyzed by mass spectroscopy. Therefore, proteomic studies allow the monitoring of regulatory changes that occur because of posttranslational modifications and quantification for large numbers of proteins simultaneously. Taking advantage of the technology used to develop DNA arrays, companies are now developing protein arrays that will be applied to new drug discovery (396).

## TRANSCRIPTIONAL CONTROL OF GASTROINTESTINAL PEPTIDES

Although knowledge in the transcriptional control of GI peptides has accelerated over the last several years, the field is still hampered by the paucity of gut-derived cell lines that express regulatory peptides. The problem has been circumvented somewhat through the use of neural and endocrine-derived hormone-producing cell lines, but application of data obtained with these models to the gut requires assumptions that may not be accurate. Future work in this field will be assisted greatly by the application of high-throughput and transgenic technologies and the development of immortalized and transformed cell lines using *in vitro* DNA transfer techniques. An overview of what has been accomplished with respect to specific GI peptides can be found primarily in Chapters 4 through 6. Nevertheless, a few peptides deserve brief mention. To date, most studies of the transcriptional control of peptide hormones have focused on somatostatin and vasoactive intestinal peptide because they are expressed in islet or neural-derived cell lines (397–400). The downside of this is that little is known about how somatostatin is regulated in gut-derived tissues; for this reason, the peptide should become a priority for future transcriptional control studies in the GI tract. Studies on the transcriptional control of gastrin have been slow for similar reasons and have been reviewed recently (401). Information on the transcriptional control of secretin and CCK has increased because of the use of transgenic mouse models (25,402–404).

## POSTTRANSCRIPTIONAL PROCESSING

### Polyadenylation

Three major events occur at the end of transcription: (1) The poly(A) tail is added, (2) adenine bases are methylated, and (3) hnRNA is processed by removing introns before exiting the nucleus (see Fig. 1-2) (405). All mRNA, except those encoding most histone proteins, have poly(A) tails. The length of the poly(A) tail that is added ranges from 200 to 250 bp and is quite uniform among eukaryotic organisms. Once the transcript reaches the cytoplasm, the length of the poly(A) sequence decreases with the age of the transcript (406). Thus, polyadenylation contributes to mRNA stability and translational activation, processes that also involve a synergistic interaction with the cap site (407–409).

Because there is no poly (dT) sequence within DNA, addition of the poly(A) tail represents a posttranscriptional modification of the newly synthesized mRNA. The AATAAA site in DNA is transcribed as AAUAAA and signals endonuclease cleavage of hnRNA ~20 bp after this RNA element (410). Several factors are required for specific recognition of the AAUAAA element before the addition of adenylate residues by poly(A) polymerase (411,412). Polyadenylation occurs in two phases: (1) an AAUAAA-dependent phase marked by addition of the first 10 residues, and (2) an AAUAAA-independent phase marked by rapid elongation and catalyzed by a poly(A)-binding protein (413). In addition, endonuclease cleavage of polyadenylated histone H1 transcripts have also been shown to require the presence of small nuclear ribonucleoproteins (U7 snRNP, pronounced "snurp"), which are trans-acting factors that participate in RNA splicing reactions (414). Transcription can proceed for up to 2 kb past the polyadenylation site and may terminate prematurely 30% of the time. Adenylate residues within exons are methylated at the sixth nitrogen and are thought to serve a protective role for those sequences that will eventually be translated (415). It is now known that formation of the PIC is linked to the assembly of factors involved in polyadenylation (416).

## RNA Splicing

### *The Spliceosome*

Soon after the termination of transcription, most vertebrate hnRNA (pre-mRNA) will be posttranscriptionally processed after exiting the nucleus into a form that can be translated (see Fig. 1-2). This involves removing intervening sequences that in some transcripts contain transcriptional regulatory signals (cis-acting elements). Splice sites are identified by comparing the genomic sequence with the cDNA prepared from an RNA template. The cis-acting elements within the intron that regulate RNA splicing are GU (GT in the genomic sequence) at the 5′ splice border, AG at the 3′ splice border, and a pyrimidine-rich element that defines the area of the branch point 20 bp upstream from the 3′ splice junction (Fig. 1-10). The branch point lies just upstream of the pyrimidine-rich region $(PyPy)_n$ and is a highly conserved sequence in yeast (UACUAAC) but much less so in vertebrates.

Five snRNA-U1, U2, U5, U4, and U6-combine with subsets of about 10 different proteins to form small nuclear ribonucleoproteins (snRNPs) (417,418). The snRNA, ranging in size from 56 to 217 nucleotides, are quite abundant in the nucleoplasm and contain a trimethylguanylate cap. Some proteins are components of all five major snRNPs, whereas others are unique to one snRNP. The U7 snRNP, which is present in low concentrations, participates in the 3′ posttranscriptional processing of hnRNA [poly(A)] (419). The five major snRNPs assemble into large multicomponent complexes called *spliceosomes* to perform the splicing reactions (420). There reactions occur in three steps: cleavage at the 5′ exon-intron border with formation of a branch point, excision of the branch point as a lariat, and joining of the exons. Splice site selection can be influenced by subtle changes in flanking exon sequences (421–423).

The basic steps in RNA processing illustrated in Figure 1-10 are as follows (419): U1 snRNP binds in a sequence-specific manner to the 5′ exon-intron junction of capped pre-mRNA (424). An U2 snRNP accessory factor (U2AF) then binds to the pyrimidine-rich element before sequence-specific recognition of the branch point element by U2 snRNP (425,426). The 5′ exon is released by cleavage of the 5′ exon junction. This allows the freed 5′ guanylate residue to form a phosphodiester bond at the 2′ site of an adenylate residue within the branch point. U4 and U6 snRNPs are paired together by complementary bases and function as a single snRNP complex (427). The recruitment of the U4/U6 snRNPs to the spliceosome is essential to the last excision step and final removal of the intron from the pre-mRNA. U4/U6 snRNP cooperates with the U2 branch point complex without direct contact with RNA (428). U5 snRNP binds just upstream of the 3′ splice junction to initiate cleavage of the 3′ intron border. Finally, the intron is removed as a lariat and the two exons are joined. More recent evidence indicates that small RNA catalyze the splicing reactions without the presence of specific enzymes (429,430). As observed for polyadenylation, the splicing events coincide with transcriptional events (431). It is therefore somewhat surprising that the events involved in splicing are not better understood. Nevertheless, with the understanding that the complexity of the human genome lies beyond the DNA sequence and at the level of epigenetics and alternative splice products, the next decade will likely witness heightened attention to this additional nuclear process (431,432).

### *Alternative Splicing*

Eukaryotic cells have applied the mechanics of RNA splicing to generate the protein diversity necessary to meet their multiple demands. Thus, in contrast with the original definition of a gene in which only one transcript is produced, complex genes can generate multiple protein isoforms from multiple RNA transcripts through alternative splicing (433). This can be achieved by altering which introns and exons are included in or excluded from the mature mRNA transcript that is used as the template for peptide chain elongation. Accordingly, the definition of introns and exons for each gene is actually a fluid concept because an intron for one gene product may become an exon within another transcript. Alternative splicing is a mechanism used by many protein classes, including muscle-related genes, hormones, and transcription factors (434–438).

### *Regulated Posttranscriptional Mechanisms*

In addition to cis-acting DNA elements, the cis and trans models of regulation also occur at the posttranscriptional level (439). Ferritin and the transferrin receptor (TfR),

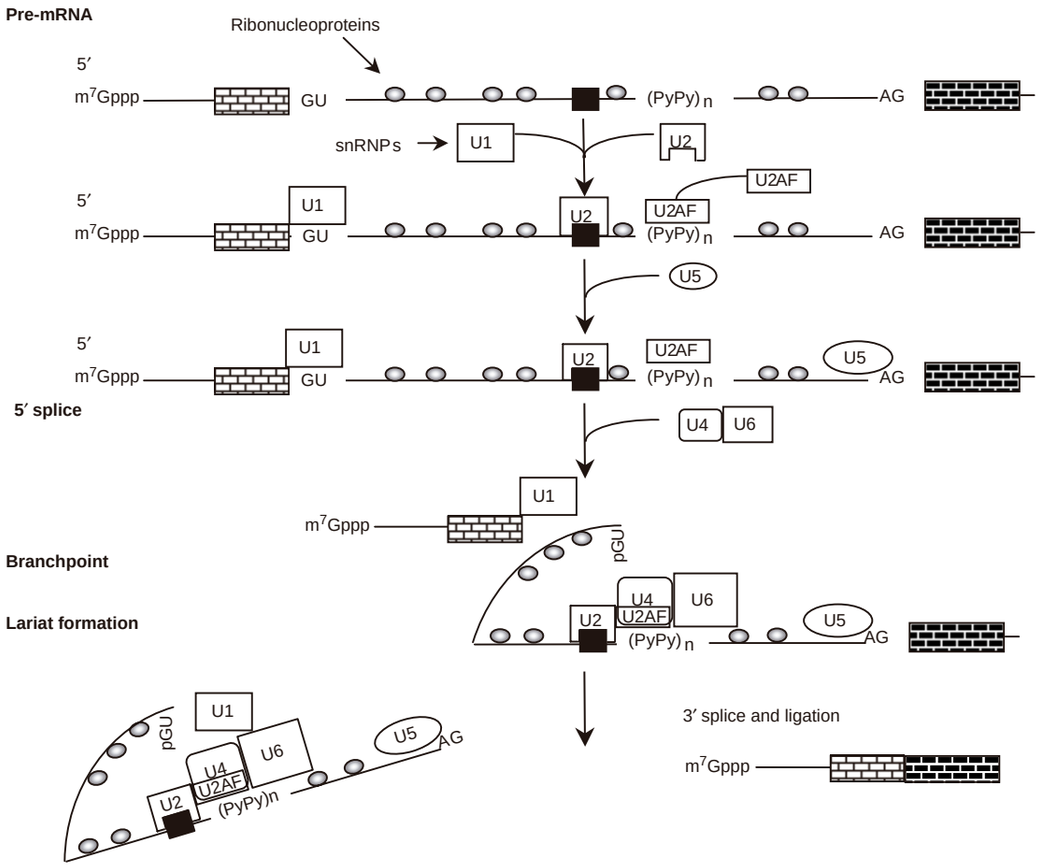



**FIG. 1-10.** Ribonucleic acid (RNA) splicing reactions. First, small ribonucleoproteins (snRNPs, pronounced "snurps") and accessory factors (U2 accessory factor [U2AF]) bind in a sequence-specific manner to the branch point and intron-exon borders. Second, the 5′ exon-intron border is cleaved, and a "lariat" is formed by the free end of the intron at the branch point. Third, the 3′ intron-exon border is cleaved, the exons are joined, and the excised intron is removed in the form of a lariat.

which regulate the storage and uptake of iron, are the best known examples of regulated posttranscriptional control (440). Cis-acting RNA elements, responsible for conferring iron regulation on both proteins (iron-response elements [IREs]), reside in the 5′ UTR and 3′ UTR of ferritin and TfR mRNA transcripts, respectively. The same iron-binding protein (IRE-BP) that binds to the IRE in the 5′ UTR of ferritin to block translation can also bind to the 3′ UTR of TfR to block mRNA degradation (439,441,442). Therefore, regulation of iron homeostasis ultimately depends on post-transcriptional mechanisms that either block translation or increase mRNA stability.

## TRANSPORT ACROSS THE NUCLEAR MEMBRANE

As noted earlier, RNA is synthesized initially as a much larger primary transcript molecule that in many instances undergoes posttranscriptional modification (e.g., splicing, degradation). However, for any mature RNA transcript to be translated, it must be transported from the nucleus to the cytoplasm. In contrast, nuclear regulatory proteins are translated in the cytoplasm and are eventually returned to the nucleus, either immediately after synthesis or after a dormant state from which they are activated in response to signals (443). This bidirectional shuttling of macromolecules between the cytoplasm and the nucleus occurs through the nuclear pore complex, a specialized compartment of the nuclear membrane regulated by a group of transport receptors called karyopherins. Both import and export processes through the nucleus require energy in the form of the Ras-related GTPase Ran and specific targeting signals on the cargo to be transported (nuclear localization and export signals) (444). The three-dimensional structure of the nuclear pore complex shows a doughnut-shaped structure comprising eight subunits (445). From the eight subunits emanate spoke-like structures that radiate inward to form a central plug (446, 447). The cytoplasmic surface of the nuclear pore complex (NPC) is closely associated with ribosomes. Its nuclear

surface is thought to participate in the organization of the genome by binding to specific DNA sequences within transcribed genes with products that may be destined for export from the nucleus (gene-gating hypothesis) (448).

## CONCLUSION

With the dawn of the postgenomic era on us, our next challenge is to apply the volumes of available genetic, molecular, and cell biological information to tackle questions of GI physiology and development. To accomplish this task and make optimal use of past, ongoing, and future discoveries, physiologists will need to acquire the basic vocabulary of several disciplines including bioinformatics. It is our hope that this chapter has laid the initial foundation necessary to understand those aspects of physiology that pertain to transcriptional control.

## ACKNOWLEDGMENTS

The work was supported by Public Health Service National Institutes of Health grants DK 55732 and DK 45729 to Dr. Merchant. The authors thank Gail Kelsey and Colleen Hill for assistance with the preparation of both the manuscript and the figures.

## REFERENCES

1. Lodish HF, Berk AJ, Zipursky LS, Matsudaira P, Baltimore D, Darnell J. *Molecular cell biology.* New York: W.H. Freeman and Co., 2000.
2. Sawadogo M, Sentenac A. RNA polymerase B (II) and general transcription factors. *Annu Rev Biochem* 1990;59:711–754.
3. Fatica A, Tollervey D. Insights into the structure and function of a guide RNP. *Nat Struct Biol* 2003;10:237–239.
4. Fitzgerald M, Shenk T. The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* 1981; 24:251–260.
5. Birnstiel M, Busslinger M, Strub K. Transcription termination and 3′ processing: the end is in site. *Cell* 1985;41:349–359.
6. McDevitt MA, Gilmartin GM, Nevins JR. Multiple factors are required for poly (A) addition to a mRNA 3′ end. *Genes Dev* 1988; 2:588–597.
7. Shatkin AJ. Capping of eukaryotic mRNAs. *Cell* 1976;9:645–654.
8. Rozen F, Sonenberg N. Identification of nuclear cap–specific proteins in HeLa cells. *Nucleic Acids Res* 1987;15:6489–6500.
9. Ausubel FM, Brent R, Moore DD, Seidman JG, Smith JA, Struhl K. *Current protocols in molecular biology*. New York: John Wiley & Sons, 2004.
10. Maniatis T, Goodbourn S, Fischer JA. Regulation of inducible and tissue-specific gene expression. *Science* 1987;236:1237–1245.
11. Chen C, Okayama H. High-efficiency transformation of mammalian cells by plasmid DNA. *Mol Cell Biol* 1987;7:2745–2752.
12. Sompayrac LM, Danna KJ. Efficient infection of monkey cells with DNA of Simian Virus 40. *Proc Natl Acad Sci U S A* 1981;78: 7575–7578.
13. Hanahan D. Transgenic mice as probes into complex systems. *Science* 1989;246:1265–1275.
14. Jaenisch R. Transgenic animals. *Science* 1988;240:1468–1474.
15. Larsson LI, Rehfeld JF. Characterization of antral gastrin cells with region-specific antisera. *J Histochem Cytochem* 1977;25:1317–1321.
16. Larsson LI, Rehfeld JF, Goltermann N. Gastrin in the human fetus. *Scand J Gastroenterol* 1977;12:869–872.
17. Powell CT, Ney C, Aran P, Agarwal K. A gastrin gene is expressed in both porcine pituitary and antral mucosal tissues. *Nucleic Acids Res* 1985;13:7299–7304.
18. Fourel G, Magdinier F, Gilson E. Insulator dynamics and the setting of chromatin domains. *Bioessays* 2004;26:523–532.
19. Zhao H, Dean A. An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. *Nucleic Acids Res* 2004;32:4903–4919.
20. Merchant JL. EGF receptor activation of the human gastrin gene: a tale of two zinc finger transcription factor families. *Keio J Med* 2000;49: 106–110.
21. Lei S, Dubeykovskiy A, Chakladar A, Wojtukiewicz L, Wang TC. The murine gastrin promoter is synergistically activated by transforming growth factor-beta/Smad and Wnt signaling pathways. *J Biol Chem* 2004;279:42492–42502.
22. Wang TC, Brand SJ. Islet cell-specific regulatory domain in the gastrin promoter contains adjacent positive and negative DNA elements. *J Biol Chem* 1990;265:8908–8914.
23. Wheeler MB, Nishitani J, Buchan AMJ, Kopin AS, Chey WY, Chang T-M, Leiter AB. Identification of a transcriptional enhancer important for enteroendocrine and pancreatic islet cell-specific expression of the secretin gene. *Mol Cell Biol* 1992;12:3531–3539.
24. Traber PG, Wu GD, Wang W. Novel DNA-binding proteins regulate intestine-specific transcription of the sucrase-isomaltase gene. *Mol Cell Biol* 1992;12:3614–3627.
25. Ray SK, Nishitani J, Petry MW, Fessing MY, Leiter AB. Novel transcriptional potentiation of BETA2/NeuroD on the secretin gene promoter by the DNA-binding protein Finb/RREB-1. *Mol Cell Biol* 2003;23:259–271.
26. Boudreau F, Zhu Y, Traber PG. Sucrase-isomaltase gene transcription requires the hepatocyte nuclear factor-1 (HNF-1) regulatory element and is regulated by the ratio of HNF-1 alpha to HNF-1 beta. *J Biol Chem* 2001;276:32122–32128.
27. Short MK, Clouthier DE, Schaefer IM, Hammer RE, Magnuson MA, Beale EG. Tissue-specific, developmental, hormonal, and dietary regulation of rat phosphoenolpyruvate carboxykinase-human growth hormone fusion genes in transgenic mice. *Mol Cell Biol* 1992;12:1007–1020.
28. Samuelson LC, Hinkle KL. Insights into the regulation of gastric acid secretion through analysis of genetically engineered mice. *Annu Rev Physiol* 2003;65:383–400.
29. Austin CP, Battey JF, Bradley A, Bucan M, Capecchi M, Collins FS, Dove WF, Duyk G, Dymecki S, Eppig JT, Grieder FB, Heintz N, Hicks G, Insel TR, Joyner A, Koller BH, Lloyd KC, Magnuson T, Moore MW, Nagy A, Pollock JD, Roses AD, Sands AT, Seed B, Skarnes WC, Snoddy J, Soriano P, Stewart DJ, Stewart F, Stillman B, Varmus H, Varticovski L, Verma IM, Vogt TF, von Melchner H, Witkowski J, Woychik RP, Wurst W, Yancopoulos GD, Young SG, Zambrowicz B. The knockout mouse project. *Nat Genet* 2004;36:921–924.
30. Sweetser DA, Birkenmeier EH, Hoppe PC, McKeel DW, Gordon JI. Mechanisms underlying generation of gradients in gene expression within the intestine: an analysis using transgenic mice containing fatty acid binding protein-human growth hormone fusion genes. *Genes Dev* 1988;2:1318–1332.
31. Roth KA, Hermiston ML, Gordon JI. Use of transgenic mice to infer the biological properties of small intestinal stem cells and to examine the lineage relationships of their descendants. *Proc Natl Acad Sci U S A* 1991;88:9407–9411.
32. Moreira PN, Giraldo P, Cozar P, Pozueta J, Jimenez A, Montoliu L, Gutierrez-Adan A. Efficient generation of transgenic mice with intact yeast artificial chromosomes by intracytoplasmic sperm injection. *Biol Reprod* 2004;71:1943–1947.
33. Krzywinski M, Wallis J, Gosele C, Bosdet I, Chiu R, Graves T, Hummel O, Layman D, Mathewson C, Wye N, Zhu B, Albracht D, Asano J, Barber S, Brown-John M, Chan S, Chand S, Cloutier A, Davito J, Fjell C, Gaige T, Ganten D, Girn N, Guggenheimer K, Himmelbauer H, Kreitler T, Leach S, Lee D, Lehrach H, Mayo M, Mead K, Olson T, Pandoh P, Prabhu AL, Shin H, Tanzer S, Thompson J, Tsai M, Walker J, Yang G, Sekhon M, Hillier L, Zimdahl H, Marziali A, Osoegawa K, Zhao S, Siddiqui A, de Jong PJ, Warren W, Mardis E, McPherson JD, Wilson R, Hubner N, Jones S, Marra M, Schein J. Integrated and sequence-ordered BAC- and YAC-based physical maps for the rat genome. *Genome Res* 2004;14:766–779.
34. Copeland NG, Jenkins NA, Court DL. Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* 2001;2:769–779.
35. Court DL, Sawitzke JA, Thomason LC. Genetic engineering using homologous recombination. *Annu Rev Genet* 2002;36:361–388.

36. Testa G, Zhang Y, Vintersten K, Benes V, Pijnappel WW, Chambers I, Smith AJ, Smith AG, Stewart AF. Engineering the mouse genome with bacterial artificial chromosomes to create multipurpose alleles. *Nat Biotechnol* 2003;21:443–447.
37. Cotta-de-Almeida V, Schonhoff S, Shibata T, Leiter A, Snapper SB. A new method for rapidly generating gene-targeting vectors by engineering BACs through homologous recombination in bacteria. *Genome Res* 2003;13:2190–2194.
38. Wu C, Morris JR. Genes, genetics, and epigenetics: a correspondence. *Science* 2001;293:1103–1105.
39. Golubovsky M, Manton KG. Genome organization and three kinds of heritable changes: general description and stochastic factors (a review). *Front Biosci* 2005;10:335–344.
40. Sarma K, Reinberg D. Histone variants meet their match. *Nat Rev Mol Cell Biol* 2005;6:139–149.
41. Isles AR, Holland AJ. Imprinted genes and mother-offspring interactions. *Early Hum Dev* 2005;81:73–77.
42. Khorasanizadeh S. The nucleosome: from genomic organization to genomic regulation. *Cell* 2004;116:259–272.
43. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 1997;389:251–260.
44. Klenova E, Ohlsson R. Poly(ADP-ribosyl)ation and epigenetics: is CTCF PARt of the plot? *Cell Cycle* 2005;4:96–101.
45. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature* 2000;403:41–45.
46. Wu J, Grunstein M. 25 years after the nucleosome model: chromatin modifications. *Trends Biochem Sci* 2000;25:619–623.
47. Lund AH, van Lohuizen M. Epigenetics and cancer. *Genes Dev* 2004; 18:2315–2335.
48. Fischle W, Wang Y, Allis CD. Binary switches and modification cassettes in histone biology and beyond. *Nature* 2003;425:475–479.
49. Wolffe A. *Chromatin: structure and function*. San Diego: Academic Press, 1998.
50. Workman JL, Kingston RE. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem* 1998; 67:545–579.
51. Sterner DE, Berger SL. Acetylation of histones and transcription-related factors. *Microbiol Mol Biol Rev* 2000;64:435–459.
52. Roth SY, Denu JM, Allis CD. Histone acetyltransferases. *Annu Rev Biochem* 2001;70:81–120.
53. Utley RT, Ikeda K, Grant PA, Cote J, Steger DJ, Eberharter A, John S, Workman JL. Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature* 1998;394:498–502.
54. Marks PA, Miller T, Richon VM. Histone deacetylases. *Curr Opin Pharmacol* 2003;3:344–351.
55. Khochbin S, Verdel A, Lemercier C, Seigneurin-Berny D. Functional significance of histone deacetylase diversity. *Curr Opin Genet Dev* 2001;11:162–166.
56. Marlett JA, McBurney MI, Slavin JL. Position of the American Dietetic Association: health implications of dietary fiber. *J Am Diet Assoc* 2002; 102:993–1000.
57. Hinnebusch BF, Meng S, Wu JT, Archer SY, Hodin RA. The effects of short-chain fatty acids on human colon cancer cell phenotype are associated with histone hyperacetylation. *J Nutr* 2002;132: 1012–1017.
58. Leder A, Leder P. Butyric acid, a potent inducer of erythroid differentiation in cultured erythroleukemic cells. *Cell* 1975;5: 319–322.
59. Riggs MG, Whittaker RG, Neumann JR, Ingram VM. n-Butyrate causes histone modification in HeLa and Friend erythroleukaemia cells. *Nature* 1977;268:462–464.
60. Sealy L, Chalkley R. The effect of sodium butyrate on histone modification. *Cell* 1978;14:115–121.
61. Kruh J. Effects of sodium butyrate, a new pharmacological agent, on cells in culture. *Mol Cell Biochem* 1982;42:65–82.
62. Candido EP, Reeves R, Davie JR. Sodium butyrate inhibits histone deacetylation in cultured cells. *Cell* 1978;14:105–113.
63. Boffa LC, Vidali G, Mann RS, Allfrey VG. Suppression of histone deacetylation in vivo and in vitro by sodium butyrate. *J Biol Chem* 1978;253:3364–3366.
64. Lutter LC, Judis L, Paretti RF. Effects of histone acetylation on chromatin topology in vivo. *Mol Cell Biol* 1992;12:5004–5014.
65. Almouzni G, Khochbin S, Dimitrov S, Wolffe AP. Histone acetylation influences both gene expression and development of Xenopus laevis. *Dev Biol* 1994;165:654–669.
66. Van Lint C, Emiliani S, Ott M, Verdin E. Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation. *EMBO J* 1996;15:1112–1120.
67. Archer SY, Hodin RA. Histone acetylation and cancer. *Curr Opin Genet Dev* 1999;9:171–174.
68. Pouillart PR. Role of butyric acid and its derivatives in the treatment of colorectal cancer and hemoglobinopathies. *Life Sci* 1998;63:1739–1760.
69. Csordas A. Butyrate, aspirin and colorectal cancer. *Eur J Cancer Prev* 1996;5:221–231.
70. Jung M. Inhibitors of histone deacetylase as new anticancer agents. *Curr Med Chem* 2001;8:1505–1511.
71. Mariadason JM, Corner GA, Augenlicht LH. Genetic reprogramming in pathways of colonic cell maturation induced by short chain fatty acids: comparison with trichostatin A, sulindac, and curcumin and implications for chemoprevention of colon cancer. *Cancer Res* 2000;60:4561–4572.
72. Van Lint C, Emiliani S, Verdin E. The expression of a small fraction of cellular genes is changed in response to histone hyperacetylation. *Gene Expr* 1996;5:245–253.
73. Kouzarides T. Acetylation: a regulatory modification to rival phosphorylation? *EMBO J* 2000;19:1176–1179.
74. Mikkelsen IM, Huseby NE, Visvikis A, Moens U. Activation of the gamma-glutamyltransferase promoter 2 in the rat colon carcinoma cell line CC531 by histone deacetylase inhibitors is mediated through the Sp1 binding motif. *Biochem Pharmacol* 2002;64:307–315.
75. Tsubaki J, Hwa V, Twigg SM, Rosenfeld RG. Differential activation of the IGF binding protein-3 promoter by butyrate in prostate cancer cells. *Endocrinology* 2002;143:1778–1788.
76. Yang J, Kawai Y, Hanson RW, Arinze IJ. Sodium butyrate induces transcription from the G alpha(i2) gene promoter through multiple Sp1 sites in the promoter and by activating the MEK-ERK signal transduction pathway. *J Biol Chem* 2001;276:25742–25752.
77. Lu Y, Lotan R. Transcriptional regulation by butyrate of mouse galectin-1 gene in embryonal carcinoma cells. *Biochim Biophys Acta* 1999;1444:85–91.
78. Taniura S, Kamitani H, Watanabe T, Eling TE. Transcriptional regulation of cyclooxygenase-1 by histone deacetylase inhibitors in normal human astrocyte cells. *J Biol Chem* 2002;277:16823–16830.
79. Kim JH, Meng S, Shei A, Hodin RA. A novel Sp1-related cis element involved in intestinal alkaline phosphatase gene transcription. *Am J Physiol* 1999;276(4 pt 1):G800–G807.
80. Sowa Y, Orita T, Hiranabe-Minamikawa S, Nakano K, Mizuno T, Nomura H, Sakai T. Histone deacetylase inhibitor activates the p21/WAF1/Cip1 gene promoter through the Sp1 sites. *Ann N Y Acad Sci* 1999;886:195–199.
81. Xiao H, Hasegawa T, Isobe K. p300 collaborates with Sp1 and Sp3 in p21(waf1/cip1) promoter activation induced by histone deacetylase inhibitor. *J Biol Chem* 2000;275:1371–1376.
82. Doetzlhofer A, Rotheneder H, Lagger G, Koranda M, Kurtev V, Brosch G, Wintersberger E, Seiser C. Histone deacetylase 1 can repress transcription by binding to Sp1. *Mol Cell Biol* 1999;19:5504–5511.
83. Maehara K, Uekawa N, Isobe K. Effects of histone acetylation on transcriptional regulation of manganese superoxide dismutase gene. *Biochem Biophys Res Commun* 2002;295:187–192.
84. Zhang Y, Dufau ML. Silencing of transcription of the human luteinizing hormone receptor gene by histone deacetylase-mSin3A complex. *J Biol Chem* 2002;277:33431–33438.
85. Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP. DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nat Genet* 2000;25: 338–342.
86. Zhu P, Martin E, Mengwasser J, Schlag P, Janssen KP, Gottlicher M. Induction of HDAC2 expression upon loss of APC in colorectal tumorigenesis. *Cancer Cell* 2004;5:455–463.
87. Ogawa H, Ishiguro K, Gaubatz S, Livingston DM, Nakatani Y. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* 2002;296:1132–1136.
88. Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat Genet* 2000;24:88–91.
89. Ma H, Baumann CT, Li H, Strahl BD, Rice R, Jelinek MA, Aswad DW, Allis CD, Hager GL, Stallcup MR. Hormone-dependent, CARM1-directed, arginine-specific methylation of histone H3 on a steroid-regulated promoter. *Curr Biol* 2001;11:1981–1985.
90. Litt MD, Simpson M, Gaszner M, Allis CD, Felsenfeld G. Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science* 2001;293:2453–2455.

91. Noma K, Allis CD, Grewal SI. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* 2001;293:1150–1155.
92. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 2001;410:120–124.
93. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, Jenuwein T. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 2000;406:593–599.
94. Yeates TO. Structures of SET domain proteins: protein lysine methyltransferases make their mark. *Cell* 2002;111:5–7.
95. Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A, Opravil S, Doyle M, Sibilia M, Jenuwein T. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* 2001;107:323–337.
96. Wang Y, Wysocka J, Sayegh J, Lee YH, Perlin JR, Leonelli L, Sonbuchner LS, McDonald CH, Cook RG, Dou Y, Roeder RG, Clarke S, Stallcup MR, Allis CD, Coonrod SA. Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science* 2004;306:279–283.
97. Cuthbert GL, Daujat S, Snowden AW, Erdjument-Bromage H, Hagiwara T, Yamada M, Schneider R, Gregory PD, Tempst P, Bannister AJ, Kouzarides T. Histone deimination antagonizes arginine methylation. *Cell* 2004;118:545–553.
98. Daujat S, Bauer UM, Shah V, Turner B, Berger S, Kouzarides T. Crosstalk between CARM1 methylation and CBP acetylation on histone H3. *Curr Biol* 2002;12:2090–2097.
99. Pal S, Yun R, Datta A, Lacomis L, Erdjument-Bromage H, Kumar J, Tempst P, Sif S. mSin3A/histone deacetylase 2- and PRMT5-containing Brg1 complex is involved in transcriptional repression of the Myc target gene cad. *Mol Cell Biol* 2003;23:7475–7487.
100. Labrador M, Corces VG. Phosphorylation of histone H3 during transcriptional activation depends on promoter structure. *Genes Dev* 2003; 17:43–48.
101. Cheung P, Allis CD, Sassone-Corsi P. Signaling to chromatin through histone modifications. *Cell* 2000;103:263–271.
102. Mahadevan LC, Willis AC, Barratt MJ. Rapid histone H3 phosphorylation in response to growth factors, phorbol esters, okadaic acid, and protein synthesis inhibitors. *Cell* 1991;65:775–783.
103. Thomson S, Clayton AL, Hazzalin CA, Rose S, Barratt MJ, Mahadevan LC. The nucleosomal response associated with immediate-early gene induction is mediated via alternative MAP kinase cascades: MSK1 as a potential histone H3/HMG-14 kinase. *EMBO J* 1999; 18:4779–4793.
104. Clayton AL, Mahadevan LC. MAP kinase-mediated phosphoacetylation of histone H3 and inducible gene regulation. *FEBS Lett* 2003; 546:51–58.
105. Keen N, Taylor S. Aurora-kinase inhibitors as anticancer agents. *Nat Rev Cancer* 2004;4:927–936.
106. Yamamoto Y, Verma UN, Prajapati S, Kwak YT, Gaynor RB. Histone H3 phosphorylation by IKK-alpha is critical for cytokine-induced gene expression. *Nature* 2003;423:655–659.
107. Soloaga A, Thomson S, Wiggin GR, Rampersaud N, Dyson MH, Hazzalin CA, Mahadevan LC, Arthur JS. MSK2 and MSK1 mediate the mitogen- and stress-induced phosphorylation of histone H3 and HMG-14. *EMBO J* 2003;22:2788–2797.
108. Ajiro K. Histone H2B phosphorylation in mammalian apoptotic cells. An association with DNA fragmentation. *J Biol Chem* 2000;275: 439–443.
109. Cheung WL, Ajiro K, Samejima K, Kloc M, Cheung P, Mizzen CA, Beeser A, Etkin LD, Chernoff J, Earnshaw WC, Allis CD. Apoptotic phosphorylation of histone H2B is mediated by mammalian sterile twenty kinase. *Cell* 2003;113:507–517.
110. Cheung P, Tanner KG, Cheung WL, Sassone-Corsi P, Denu JM, Allis CD. Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Mol Cell* 2000;5:905–915.
111. Lo WS, Trievel RC, Rojas JR, Duggan L, Hsu JY, Allis CD, Marmorstein R, Berger SL. Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. *Mol Cell* 2000;5:917–926.
112. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 1999;99:247–257.
113. Sado T, Fenner MH, Tan SS, Tam P, Shioda T, Li E. X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev Biol* 2000;225: 294–303.
114. Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, Gartler SM. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc Natl Acad Sci U S A* 1999;96:14412–14417.
115. Antequera F, Bird A. Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A* 1993;90:11995–11999.
116. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 2002;3:415–428.
117. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* 2002;16:6–21.
118. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 1997;13:335–340.
119. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* 2002;99: 3740–3745.
120. Lee JH, Park SJ, Abraham SC, Seo JS, Nam JH, Choi C, Juhng SW, Rashid A, Hamilton SR, Wu TT. Frequent CpG island methylation in precursor lesions and early gastric adenocarcinomas. *Oncogene* 2004; 23:4646–4654.
121. Suter CM, Martin DI, Ward RL. Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet* 2004;36:497–501.
122. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 1983; 301:89–92.
123. Huang S. Histone methyltransferases, diet nutrients and tumour suppressors. *Nat Rev Cancer* 2002;2:469–476.
124. Iwama A, Oguro H, Negishi M, Kato Y, Morita Y, Tsukui H, Ema H, Kamijo T, Katoh-Fukui Y, Koseki H, van Lohuizen M, Nakauchi H. Enhanced self-renewal of hematopoietic stem cells mediated by the polycomb gene product Bmi-1. *Immunity* 2004;21:843–851.
125. Lessard J, Sauvageau G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* 2003;423:255–260.
126. Park IK, Qian D, Kiel M, Becker MW, Pihalja M, Weissman IL, Morrison SJ, Clarke MF. Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* 2003;423:302–305.
127. Milne TA, Briggs SD, Brock HW, Martin ME, Gibbs D, Allis CD, Hess JL. MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol Cell* 2002;10:1107–1117.
128. Ernst P, Wang J, Huang M, Goodman RH, Korsmeyer SJ. MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. *Mol Cell Biol* 2001;21:2249–2258.
129. Rozenblatt-Rosen O, Rozovskaia T, Burakov D, Sedkov Y, Tillib S, Blechman J, Nakamura T, Croce CM, Mazo A, Canaani E. The C-terminal SET domains of ALL-1 and TRITHORAX interact with the INI1 and SNR1 proteins, components of the SWI/SNF complex. *Proc Natl Acad Sci U S A* 1998;95:4152–4157.
130. Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, Yagyu R, Nakamura Y. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat Cell Biol* 2004;6:731–740.
131. Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 1998;19:187–191.
132. Hansen RS, Gartler SM, Scott CR, Chen SH, Laird CD. Methylation analysis of CGG sites in the CpG island of the human FMR1 gene. *Hum Mol Genet* 1992;1:571–578.
133. Murata T, Kurokawa R, Krones A, Tatsumi K, Ishii M, Taki T, Masuno M, Ohashi H, Yanagisawa M, Rosenfeld MG, Glass CK, Hayashi Y. Defect of histone acetyltransferase activity of the nuclear transcriptional coactivator CBP in Rubinstein-Taybi syndrome. *Hum Mol Genet* 2001;10:1071–1076.
134. Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, Goldman H, Jessup JM, Kolodner R. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res* 1997; 57:808–811.
135. Ogawa O, Eccles MR, Szeto J, McNoe LA, Yun K, Maw MA, Smith PJ, Reeve AE. Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour. *Nature* 1993;362:749–751.
136. Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, Bronson RT, Li E, Livingston DM, Eckner R. Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* 1998;93:361–372.

137. Xu W, Edmondson DG, Evrard YA, Wakamiya M, Behringer RR, Roth SY. Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. *Nat Genet* 2000;26:229–232.
138. Gayther SA, Batley SJ, Linger L, Bannister A, Thorpe K, Chin SF, Daigo Y, Russell P, Wilson A, Sowter HM, Delhanty JD, Ponder BA, Kouzarides T, Caldas C. Mutations truncating the EP300 acetylase in human cancers. *Nat Genet* 2000;24:300–303.
139. Kouzarides T. Histone acetylases and deacetylases in cell proliferation. *Curr Opin Genet Dev* 1999;9:40–48.
140. Borrow J, Stanton VP Jr, Andresen JM, Becher R, Behm FG, Chaganti RS, Civin CI, Disteche C, Dube I, Frischauf AM, Horsman D, Mitelman F, Volinia S, Watmore AE, Housman DE. The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. *Nat Genet* 1996;14:33–41.
141. Carapeti M, Aguiar RC, Watmore AE, Goldman JM, Cross NC. Consistent fusion of MOZ and TIF2 in AML with inv(8)(p11q13). *Cancer Genet Cytogenet* 1999;113:70–72.
142. Redner RL, Wang J, Liu JM. Chromatin remodeling and leukemia: new therapeutic paradigms. *Blood* 1999;94:417–428.
143. Grignani F, De Matteis S, Nervi C, Tomassoni L, Gelmetti V, Cioce M, Fanelli M, Ruthardt M, Ferrara FF, Zamir I, Seiser C, Lazar MA, Minucci S, Pelicci PG. Fusion proteins of the retinoic acid receptor-alpha recruit histone deacetylase in promyelocytic leukaemia. *Nature* 1998;391:815–818.
144. Jones LK, Saha V. Chromatin modification, leukaemia and implications for therapy. *Br J Haematol* 2002;118:714–727.
145. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 2004;429:457–463.
146. Epstein R. *Human molecular biology*. Cambridge: Cambridge University Press, 2003.
147. Dynan WS, Tjian R. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* 1983; 35:79–87.
148. Van Dyke MW, Roeder RG, Sawadogo M. Physical analysis of transcription preinitiation complex assembly on a Class II gene promoter. *Science* 1988;241:1335–1338.
149. Hai T, Horikoshi M, Roeder RG, Green MR. Analysis of the role of the transcription factor ATF in the assembly of a functional preinitiation complex. *Cell* 1988;54:1043–1051.
150. Dynan WS, Tjian R. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature* 1985;316:774–778.
151. McKnight S, Tjian R. Transcriptional selectivity of viral genes in mammalian cells. *Cell* 1986;46:795–805.
152. Mitchell PJ, Tjian R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 1989;245: 371–378.
153. Dynan WS. Modularity in promoters and enhancers. *Cell* 1989;58:1–4.
154. Brown DD. The role of stable complexes that repress and activate eucaryotic genes. *Cell* 1984;37:359–365.
155. Renkawitz R. Transcriptional repression in eukaryotes. *Trends Genet* 1990;6:192–197.
156. Jackson SM, Keech CA, Williamson DJ, Gutierrez-Hartmann A. Interaction of basal positive and negative transcription elements controls repression of the proximal rat prolactin promoter in nonpituitary cells. *Mol Cell Biol* 1992;12:2708–2719.
157. Hampsey M. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev* 1998;62:465–503.
158. Struhl K. Yeast transcriptional regulatory mechanisms. *Annu Rev Genet* 1995;29:651–674.
159. Hahn S, Buratowski S, Sharp PA, Guarente L. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proc Natl Acad Sci U S A* 1989; 86:5718–5722.
160. Pugh BF, Tjian R. Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* 1990;61:1187–1197.
161. Wefald FC, Devlin BH, Williams RS. Functional heterogeneity of mammalian TATA-box sequences revealed by interaction with a cell-specific enhancer. *Nature* 1990;344:260–262.
162. Sawadogo M, Roeder RG. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* 1985;43:165–175.
163. Horikoshi M, Hai T, Lin Y-S, Green MR, Roeder RG. Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. *Cell* 1988;54:1033–1042.
164. Nakajima N, Horikoshi M, Roeder RG. Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID. *Mol Cell Biol* 1988;8:4028–4040.
165. Burke TW, Kadonaga JT. Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev* 1996;10:711–724.
166. Smale ST, Baltimore D. The "initiator" as a transcription control element. *Cell* 1989;57:103–113.
167. Pugh BF, Tjian R. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev* 1991;5:1935–1945.
168. Weis L, Reinberg D. Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *FASEB J* 1992;6:3300–3309.
169. Smale ST, Schmidt MC, Berk AJ, Baltimore D. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. *Proc Natl Acad Sci U S A* 1990;87:4509–4513.
170. Wiborg O, Berglund L, Boel E, Norris F, Norris K, Rehfeld JF, Marcker KA, Vuust J. Structure of a human gastrin gene. *Proc Natl Acad Sci U S A* 1984;81:1067–1069.
171. Andrisani OM, Dixon JE. Somatostatin gene regulation. *Annu Rev Physiol* 1990;52:793–806.
172. Haun RS, Dixon JE. A transcriptional enhancer essential for the expression of the rat cholecystokinin gene contains a sequence identical to the -296 element of the human c-fos gene. *J Biol Chem* 1990;265:15455–15463.
173. Philippe J, Drucker DJ, Knepel W, Jepeal L, Misulovin Z, Habener JF. Alpha-cell-specific expression of the glucagon gene is conferred to the glucagon promoter element by the interactions of DNA-binding proteins. *Mol Cell Biol* 1988;8:4877–4888.
174. Kopin AS, Wheeler MB, Nishitani J, McBride EW, Chang T-M, Chey WY, Leiter AB. The secretin gene: evolutionary history, alternative splicing, and developmental regulation. *Proc Natl Acad Sci U S A* 1991; 88:5335–5339.
175. Blasband AJ, Rogers KT, Chen X, Azizkhan JC, Lee DC. Characterization of the rat transforming growth factor a gene and identification of promoter sequences. *Mol Cell Biol* 1990;10:2111–2121.
176. Guarente L, Hoar E. Upstream activation sites of the CYC1 gene of Saccharomyces cerevisiae are active when inverted but not when placed downstream of the "TATA box." *Proc Natl Acad Sci U S A* 1984;81:7860–7864.
177. Struhl K. Genetic properties and chromatin structure of the yeast gal regulatory element: an enhancer-like sequence. *Proc Natl Acad Sci U S A* 1984;81:7865–7869.
178. Winter E, Varshavsky A. A DNA binding protein that recognizes oligo(dA).oligo(dT) tracts. *EMBO J* 1989;8:1867–1877.
179. Koch KA, Thiele DJ. Functional analysis of a homopolymeric (dA-dT) element that provides nucleosomal access to yeast and mammalian transcription factors. *J Biol Chem* 1999;274:23752–23760.
180. Sharp PA. TATA-binding protein is a classless factor. *Cell* 1992; 68:819–821.
181. Buratowski S, Hahn S, Guarente L, Sharp PA. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* 1989; 56:549–561.
182. Greenblatt J. Riding high on the TATA box. *Nature* 1992;360:16–17.
183. Nikolov DB, Hu S-H, Lin J, Gasch A, Hoffmann A, Horikoshi M, Chua N-H, Roeder RG, Burley SK. Crystal structure of TFIID TATA-box binding protein. *Nature* 1992;360:40–46.
184. Hoey T, Weinzierl ROJ, Gill G, Chen J-L, Dynlacht BD, Tjian R. Molecular cloning and functional analysis of Drosophila TAF110 reveal properties expected of coactivators. *Cell* 1993;72:247–260.
185. Green MR. TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem Sci* 2000; 25:59–63.
186. Wassarman DA, Sauer F. TAF(II)250: a transcription toolbox. *J Cell Sci* 2001;114(pt 16):2895–2902.
187. Comai L, Tanes N, Tjian R. The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. *Cell* 1992;68:965–976.
188. Rigby PWJ. Three in one and one in three: it all depends on TBP. *Cell* 1993;72:7–10.
189. Bell SP, Pikaard CS, Reeder RH, Tjian R. Molecular mechanisms governing species-specific transcription of ribosomal RNA. *Cell* 1989; 59:489–497.
190. Taggart AKP, Fisher TS, Pugh BF. The TATA-binding protein and associated factors are components of Pol III transcription factor TFIIIB. *Cell* 1992;71:1015–1028.

191. Schultz MC, Reeder RH, Hahn S. Variants of the TATA-binding protein can distinguish subsets of RNA Polymerase I, II, and III promoters. *Cell* 1992;69:697–702.
192. Peterson MG, Tjian R. The tell-tail trigger. *Nature* 1992;358:620–621.
193. Fischer L, Gerard M, Chalut C, Lutz Y, Humbert S, Kanno M, Chambon P, Egly J-M. Cloning of the 62-kilodalton component of basic transcription factor BTF2. *Science* 1992;257:1392–1395.
194. Lu H, Zawel L, Fisher L, Egly J-M, Reinberg D. Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* 1992;358:641–645.
195. Feaver WJ, Svejstrup JQ, Henry NL, Kornberg RD. Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. *Cell* 1994;79:1103–1109.
196. Sopta M, Burton ZF, Greenblatt J. Structure and associated DNA-helicase activity of a general transcription initiation factor that binds to RNA polymerase II. *Nature* 1989;341:410–414.
197. Roy AL, Meisterernst M, Pognonec P, Roeder RG. Cooperative interaction of an initiator-binding transcripton initiation factor and the helix-loop-helix activator USF. *Nature* 1991;354:245–248.
198. Liao SM, Zhang J, Jeffery DA, Koleske AJ, Thompson CM, Chao DM, Viljoen M, van Vuuren HJ, Young RA. A kinase-cyclin pair in the RNA polymerase II holoenzyme. *Nature* 1995;374:193–196.
199. Cedar H. DNA methylation and gene activity. *Cell* 1988;53:3–4.
200. Gross DS, Garrard WT. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem* 1988;57:159–197.
201. Bird A. The essentials of DNA methylation. *Cell* 1992;70:5–8.
202. Levine M, Manley JL. Transcriptional repression of eukaryotic promoters. *Cell* 1989;59:405–408.
203. Ponta H, Cato ACB, Herrlich P. Interference of pathway specific transcription factors. *Biochim Biophys Acta* 1992;1129:255–261.
204. Diamond MI, Miner JN, Yoshinaga SK, Yamamoto KR. Transcription factor interactions: selectors of positive or negative regulation from a single DNA element. *Science* 1990;249:1266–1272.
205. Belayew A, Tilghman SM. Genetic analysis of a-fetoprotein synthesis in mice. *Mol Cell Biol* 1982;2:1427–1435.
206. Watanabe A, Miyazaki M, Taketa K. Prompt elevation of rat serum alpha-fetoprotein by acute liver injury following a single injection of ethionine. *Int J Cancer* 1976;17:518–524.
207. Benchimol S, Fuks A, Jothy S, Beauchemin N, Shirota K, Stanners CP. Carcinoembryonic antigen, a human tumor marker, functions as an intercellular adhesion molecule. *Cell* 1989;57:327–334.
208. Brand AH, Breeden L, Abraham J, Sternglanz R, Nasmyth N. Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* 1985;41:41–48.
209. Cao SX, Gutman PD, Dave HPG, Schechter AN. Identification of a transcriptional silencer in the 5′-flanking region of the human epsilon-globin gene. *Proc Natl Acad Sci U S A* 1989;86:5306–5309.
210. Pierce JW, Gifford AM, Baltimore D. Silencing of the expression of the immunoglobulin k gene in non-B cells. *Mol Cell Biol* 1991; 11:1431–1437.
211. Jones NC, Rigby PWJ, Ziff EB. Trans-acting protein factors and the regulation of eukaryotic transcription: lessons from studies on DNA tumor viruses. *Genes Dev* 1988;2:267–281.
212. Schleif R. DNA binding by proteins. *Science* 1988;241:1182–1187.
213. Berg OG, von Hippel PH. Selection of DNA binding sites by regulatory proteins. *Trends Biochem Sci* 1988;13:207–211.
214. Johnson PF, McKnight SL. Eukaryotic transcriptional regulatory proteins. *Annu Rev Biochem* 1989;58:799–839.
215. Harrison SC. A structural taxonomy of DNA-binding domains. *Nature* 1991;353:715–719.
216. Hermanson O, Glass CK, Rosenfeld MG. Nuclear receptor coregulators: multiple modes of modification. *Trends Endocrinol Metab* 2002;13:55–60.
217. Levin M, Hoey T. Homeobox proteins as sequence-specific transcription factors. *Cell* 1988;55:537–540.
218. Wright CVE, Cho KWY, Oliver G, DeRobertis EM. Vertebrate homeodomain proteins: families of region-specific transcription factors. *Trends Biochem Sci* 1989;14:52–56.
219. Otting G, Qian Y, Muller M, Affolter M, Gehring WJ, Wuthrich K. Secondary structure determination for the Antennapedia homeodomain by nuclear magnetic resonance: evidence for a helix-turn-helix motif. *EMBO J* 1988;7:4305–4309.
220. Marx JL. Homeobox linked to gene control. *Science* 1988;242: 1008–1009.
221. Ingraham HA, Chen R, Mangalam HJ, Elsholtz HP, Flynn SE, Lin CR, Simmons DM, Swanson L, Rosenfeld MG. A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell* 1988;55:519–529.
222. Bodner M, Castrillo J-L, Theill LE, Deerinck T, Ellisman M, Karin M. The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. *Cell* 1988;55:505–518.
223. Baumruker T, Sturm R, Herr W. OBP100 binds remarkably degenerate octamer motifs through specific interactions with flanking sequences. *Genes Dev* 1988;2:1400–1413.
224. LeBowitz JH, Kobayashi T, Staudt L, Baltimore D, Sharp PA. Octamer-binding proteins from B or HeLA cells stimulate transcription of the immunoglobulin heavy-chain promoter in vitro. *Genes Dev* 1988; 2:1227–1237.
225. Stern S, Tanaka M, Herr W. The Oct-1 homoeodomain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. *Nature* 1989;341:624–630.
226. Holland PWH, Hogan BLM. Expression of homeobox genes during mouse development: a review. *Genes Dev* 1988;2:773–782.
227. Wolgemuth DJ, Behringer RR, Mostoller MP, Brinster RL, Palmiter RD. Transgenic mice overexpressing the mouse homeobox-containing gene Hox-1.4 exhibit abnormal gut development. *Nature* 1989;337: 464–467.
228. Duprey P, Chowdhury K, Dressler GR, Balling R, Simon D, Guenet J-L, Gruss P. A mouse gene homologous to the Drosophila gene caudal is expressed in epithelial cells from the embryonic intestine. *Genes Dev* 1988;2:1647–1654.
229. Beck F. The role of Cdx genes in the mammalian gut. *Gut* 2004; 53:1394–1396.
230. Melloul D. Transcription factors in islet development and physiology: role of PDX-1 in beta-cell function. *Ann N Y Acad Sci* 2004;1014:28–37.
231. Burke AC, Nelson CE, Morgan BA, Tabin C. Hox genes and the evolution of vertebrate axial morphology. *Development* 1995;121:333–346.
232. Brooke NM, Garcia-Fernandez J, Holland PW. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* 1998; 392:920–922.
233. Jonsson J, Carlsson L, Edlund T, Edlund H. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* 1994;371:606–609.
234. Offield MF, Jetton TL, Labosky PA, Ray M, Stein RW, Magnuson MA, Hogan BL, Wright CV. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* 1996;122: 983–995.
235. James R, Kazenwadel J. Homeobox gene expression in the intestinal epithelium of adult mice. *J Biol Chem* 1991;266:3246–3251.
236. Li MK, Folpe AL. CDX-2, a new marker for adenocarcinoma of gastrointestinal origin. *Adv Anat Pathol* 2004;11:101–105.
237. Mutoh H, Sakurai S, Satoh K, Tamada K, Kita H, Osawa H, Tomiyama T, Sato Y, Yamamoto H, Isoda N, Yoshida T, Ido K, Sugano K. Development of gastric carcinoma from intestinal metaplasia in Cdx2-transgenic mice. *Cancer Res* 2004;64:7740–7747.
238. Katoh M. Human FOX gene family (Review). *Int J Oncol* 2004; 25:1495–1500.
239. Cillo C, Faiella A, Cantile M, Boncinelli E. Homeobox genes and cancer. *Exp Cell Res* 1999;248:1–9.
240. Hromas R, Costa R. The hepatocyte nuclear factor-3/forkhead transcription regulatory family in development, inflammation, and neoplasia. *Crit Rev Oncol Hematol* 1995;20(1-2):129–140.
241. Berg JM. Potential metal-binding domains in nucleic acid binding proteins. *Science* 1986;232:485–487.
242. Berg JM. Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc Natl Acad Sci U S A* 1988;85:99–102.
243. Miller J, McLachlan AD, Klug A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J* 1985;4:1609–1614.
244. Schwabe JWR, Rhodes D. Beyond zinc fingers: steroid hormone receptors have a novel structural motif for DNA recognition. *Trends Biochem Sci* 1991;16:291–296.
245. Evans RM, Hollenberg SM. Zinc fingers: gilt by association. *Cell* 1988; 52:1–3.
246. Coleman JE. ZINC PROTEINS: enzymes, storage proteins, transcription factors, and replication proteins. *Annu Rev Biochem* 1992; 61:897–946.
247. Luisi BF, Xu WX, Otwinowski Z, Freedman LP, Yamamoto KR, Sigler PB. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 1991;352:497–505.
248. Pavletich NP, Pabo CO. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 1991;252:809–817.

249. Marmorstein R, Carey M, Ptashne M, Harrison SC. DNA recognition by GAL4: structure of a protein-DNA complex. *Nature* 1992;356: 408–414.
250. Luisi B. Zinc standard for economy. *Nature* 1992;356:379–380.
251. Jordan SR, Pabo CO. Structure of the lambda complex at 2.5 A resolution: details of the repressor-operator interactions. *Science* 1988; 242:893–899.
252. Aggarwal AK, Rodgers DW, Drottar M, Ptashne M, Harrison SC. Recognition of a DNA operator by the repressor phage 434: a view at high resolution. *Science* 1988;242:899–907.
253. Kissinger CR, Liu B, Martin-Blanco E, Kornberg TB, Pabo C. Crystal structure of an engrailed Homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. *Cell* 1990;63:579–590.
254. O'Neil KT, Hoess RH, DeGrado WF. Design of DNA-binding peptides based on the leucine zipper motif. *Science* 1990;249:774–778.
255. Falke D, Juliano RL. Selective gene regulation with designed transcription factors: implications for therapy. *Curr Opin Mol Ther* 2003; 5:161–166.
256. Jouvenot Y, Ginjala V, Zhang L, Liu PQ, Oshimura M, Feinberg AP, Wolffe AP, Ohlsson R, Gregory PD. Targeted regulation of imprinted genes by synthetic zinc-finger transcription factors. *Gene Ther* 2003; 10:513–522.
257. Moore M, Ullman C. Recent developments in the engineering of zinc finger proteins. *Brief Funct Genomic Proteomic* 2003;1:342–355.
258. Landschulz WH, Johnson PF, McKnight SL. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* 1988;240:1759–1764.
259. Gentz R, Rauscher FJ III, Abate C, Curran T. Parallel association of fos and jun leucine zippers juxtaposes DNA binding domains. *Science* 1989;243:1695–1699.
260. Pu WT, Struhl K. The leucine zipper symmetrically positions the adjacent basic regions for specific DNA binding. *Proc Natl Acad Sci U S A* 1991;88:6901–6905.
261. Cuenoud B, Schepartz A. Altered specificity of DNA-binding proteins with transition metal dimerization domains. *Science* 1993; 259:510–513.
262. O'Shea EK, Rutkowski R, Kim PS. Evidence that the leucine zipper is a coiled coil. *Science* 1989;243:538–542.
263. Yun Y, Dumoulin M, Habener JF. DNA-binding and dimerization domains of adenosine 3′,5′-cyclic monophosphate-responsive protein CREB reside in the carboxyl-terminal 66 amino acids. *Mol Endocrinol* 1990;4:931–939.
264. Landschulz WH, Johnson PF, McKnight SL. The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. *Science* 1989;243:1681–1688.
265. Turner R, Tjian R. Leucine repeats and an adjacent DNA binding domain mediate the formation of functional cFos-cJun heterodimers. *Science* 1989;243:1689–1694.
266. Ellenberger TE, Brandl CJ, Struhl K, Harrison SC. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell* 1992; 71:1223–1237.
267. Murre C, McCaw PS, Baltimore D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 1989;56:777–783.
268. Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka SD, Lassar AB, Weintraub H, Baltimore D. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* 1989;58:537–544.
269. Williams T, Tjian R. Characterization of the dimerization motif in AP-2 and its function in heterologous DNA-binding proteins. *Science* 1991; 251:1067–1071.
270. Habener JF. Cyclic AMP response element binding proteins: a cornucopia of transcription factors. *Mol Endocrinol* 1990;4:1087–1094.
271. Ziff EB. Transcription factors: a new family gathers at the cAMP response site. *Trends Genet* 1990;6:69–72.
272. Hai T, Curran T. Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. *Proc Natl Acad Sci U S A* 1991;88:3720–3724.
273. Blackwell TK, Weintraub H. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* 1990;250:1104–1110.
274. Moser M, Imhof A, Pscherer A, Bauer R, Amselgruber W, Sinowatz F, Hofstadter F, Schule R, Buettner R. Cloning and characterization of a second AP-2 transcription factor: AP-2 beta. *Development* 1995; 121:2779–2788.
275. Ohtaka-Maruyama C, Wang X, Ge H, Chepelinsky AB. Overlapping Sp1 and AP2 binding sites in a promoter element of the lens-specific MIP gene. *Nucleic Acids Res* 1998;26:407–414.
276. Benezra R, Davis RL, Lockshon D, Turner DL, Weintraub H. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell* 1990;61:49–59.
277. Hai T, Hartman MG. The molecular biology and nomenclature of the activating transcription factor/cAMP responsive element binding family of transcription factors: activating transcription factor proteins and homeostasis. *Gene* 2001;273:1–11.
278. Cowell IG, Skinner A, Hurst HC. Transcriptional repression by a novel member of the bZIP family of transcription factors. *Mol Cell Biol* 1992;12:3070–3077.
279. Ptashne M, Gann AAF. Activators and targets. *Nature* 1990;346: 329–331.
280. Ptashne M. How eukaryotic transcriptional activators work. *Nature* 1988;335:683–689.
281. Kadonaga JT, Courey AJ, Ladika J, Tjian R. Distinct regions of Sp1 modulate DNA binding and transcriptional activation. *Science* 1988; 242:1566–1570.
282. Courey AJ, Tjian R. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* 1988;55:887–898.
283. Courey AJ, Holtzman DA, Jackson SP, Tjian R. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell* 1989;59:827–836.
284. Mermod N, O'Neill EA, Kelly TJ, Tjian R. The proline-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA binding domain. *Cell* 1989;58:741–753.
285. Moustakas A, Kardassis D. Regulation of the human p21/WAF1/Cip1 promoter in hepatic cells by functional interactions between Sp1 and Smad family members. *Proc Natl Acad Sci U S A* 1998;95:6733–6738.
286. Kardassis D, Papakosta P, Pardali K, Moustakas A. c-Jun transactivates the promoter of the human p21(WAF1/Cip1) gene by acting as a superactivator of the ubiquitous transcription factor Sp1. *J Biol Chem* 1999;274:29572–29581.
287. Chen BK, Chang WC. Functional interaction between c-Jun and promoter factor Sp1 in epidermal growth factor-induced gene expression of human 12(S)-lipoxygenase. *Proc Natl Acad Sci U S A* 2000; 97:10406–10411.
288. Warner DR, Roberts EA, Greene RM, Pisano MM. Identification of novel Smad binding proteins. *Biochem Biophys Res Commun* 2003; 312:1185–1190.
289. ten Dijke P, Hill CS. New insights into TGF-beta-Smad signalling. *Trends Biochem Sci* 2004;29:265–273.
290. Hunter T, Karin M. The regulation of transcription by phosphorylation. *Cell* 1992;70:375–387.
291. Taylor SS. cAMP-dependent protein kinase. *J Biol Chem* 1989;264: 8443–8446.
292. Roesler WJ, Vandenbark GR, Hanson RW. Cyclic AMP and the induction of eukaryotic gene transcription. *J Biol Chem* 1988;263:9063–9066.
293. Nishizuka Y. Intracellular signaling by hydrolysis of phospholipids and activation of protein kinase C. *Science* 1992;258:607–614.
294. Hunter T. A thousand and one protein kinases. *Cell* 1987;50:823–829.
295. Lin A, Frost J, Deng T, Smeal T, Al-Alawi N, Kikkawa U, Hunter T, Brenner D, Karin M. Casein kinase II is a negative regulator of c-jun DNA binding and AP-1 activity. *Cell* 1992;70:777–789.
296. Jackson SP, MacDonald JJ, Lees-Miller S, Tjian R. GC box binding induces phosphorylation of Sp1 by a DNA-dependent protein kinase. *Cell* 1990;63:155–165.
297. Gottlieb TM, Jackson SP. The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* 1993; 72:131–142.
298. Shirakawa F, Mizel SB. In vitro activation and nuclear translocation of NF-κB catalyzed by cyclic AMP-dependent protein kinase and protein kinase C. *Mol Cell Biol* 1989;9:2424–2430.
299. Lenardo MJ, Baltimore D. NF-κB: a pleiotropic mediator of inducible and tissue-specific gene control. *Cell* 1989;58:227–229.
300. Baeuerle PA, Baltimore D. I kB: A specific inhibitor of the NF-κB transcription factor. *Science* 1988;242:540–546.
301. Ravid T, Hochstrasser M. NF-kappaB signaling: flipping the switch with polyubiquitin chains. *Curr Biol* 2004;14:R898–R900.
302. Hayden MS, Ghosh S. Signaling to NF-kappaB. *Genes Dev* 2004; 18:2195–2224.

303. Clevers H. At the crossroads of inflammation and cancer. *Cell* 2004; 118:671–674.
304. Banninger G, Reich NC. STAT2 nuclear trafficking. *J Biol Chem* 2004;279:39199–39206.
305. Vinkemeier U. Getting the message across, STAT! Design principles of a molecular signaling circuit. *J Cell Biol* 2004;167:197–201.
306. Macian F, Lopez-Rodriguez C, Rao A. Partners in transcription: NFAT and AP-1. *Oncogene* 2001;20:2476–2489.
307. Tamrakar S, Rubin E, Ludlow JW. Role of pRB dephosphorylation in cell cycle regulation. *Front Biosci* 2000;5:D121–D137.
308. Prywes R, Roeder RG. Inducible binding of a factor to the c-fos enhancer. *Cell* 1986;47:777–784.
309. Norman C, Runswick M, Pollock R, Treisman R. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* 1988;55:989–1003.
310. Janknecht R, Hipskind RA, Houthaeve T, Nordheim A, Stunnenberg HG. Identification of multiple SRF N-terminal phosphorylation sites affecting DNA binding properties. *EMBO J* 1992;11:1045–1054.
311. Gauthier-Rouviere C, Basset M, Blanchard J-M, Cavadore J-C, Fernandez A, Lamb NJC. Casein kinase II induces c-fos expression via the serum response element pathway and p67srf phosphorylation in living fibroblasts. *EMBO J* 1991;10:2921–2930.
312. Jackson SP, Tjian R. O-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. *Cell* 1988;55:125–133.
313. Jackson SP, Tjian R. Purification and analysis or RNA polymerase II transcription factors by using wheat germ agglutinin affinity chromatography. *Proc Natl Acad Sci U S A* 1989;86:1781–1785.
314. Brasse-Lagnel C, Fairand A, Lavoinne A, Husson A. Glutamine stimulates argininosuccinate synthetase gene expression through cytosolic O-glycosylation of Sp1 in Caco-2 cells. *J Biol Chem* 2003;278: 52504–52510.
315. Kang HT, Ju JW, Cho JW, Hwang ES. Down-regulation of Sp1 activity through modulation of O-glycosylation by treatment with a low glucose mimetic, 2-deoxyglucose. *J Biol Chem* 2003;278:51223–51231.
316. Gewinner C, Hart G, Zachara N, Cole R, Beisenherz-Huss C, Groner B. The coactivator of transcription CREB-binding protein interacts preferentially with the glycosylated form of Stat5. *J Biol Chem* 2004; 279:3563–3572.
317. Meier CA. Co-activators and co-repressors: mediators of gene activation by nuclear hormone receptors. *Eur J Endocrinol* 1996; 134:158–159.
318. Shibata H, Spencer TE, Onate SA, Jenster G, Tsai SY, Tsai MJ, O'Malley BW. Role of co-activators and co-repressors in the mechanism of steroid/thyroid receptor action. *Recent Prog Horm Res* 1997; 52:141–165.
319. Goldman PS, Tran VK, Goodman RH. The multifunctional role of the co-activator CBP in transcriptional regulation. *Recent Prog Horm Res* 1997;52:103–120.
320. Giordano A, Avantaggiati ML. p300 and CBP: partners for life and death. *J Cell Physiol* 1999;181:218–230.
321. Roeder RG. Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett* 2005;579:909–915.
322. Kumar R, Wang RA, Barnes CJ. Coregulators and chromatin remodeling in transcriptional control. *Mol Carcinog* 2004;41:221–230.
323. Calo V, Migliavacca M, Bazan V, Macaluso M, Buscemi M, Gebbia N, Russo A. STAT proteins: from normal control of cellular events to tumorigenesis. *J Cell Physiol* 2003;197:157–168.
324. Kemper JK, Kim H, Miao J, Bhalla S, Bae Y. Role of an mSin3A-Swi/Snf chromatin remodeling complex in the feedback repression of bile acid biosynthesis by SHP. *Mol Cell Biol* 2004;24:7707–7719.
325. Moehren U, Dressel U, Reeb CA, Vaisanen S, Dunlop TW, Carlberg C, Baniahmad A. The highly conserved region of the co-repressor Sin3A functionally interacts with the co-repressor Alien. *Nucleic Acids Res* 2004;32:2995–3004.
326. Ellenrieder V, Zhang JS, Kaczynski J, Urrutia R. Signaling disrupts mSin3A binding to the Mad1-like Sin3-interacting domain of TIEG2, an Sp1-like repressor. *EMBO J* 2002;21:2451–2460.
327. O'Shea JJ, Gadina M, Schreiber RD. Cytokine signaling in 2002: new surprises in the Jak/Stat pathway. *Cell* 2002;109(suppl):S121–S131.
328. Schmidt D, Muller S. PIAS/SUMO: new partners in transcriptional regulation. *Cell Mol Life Sci* 2003;60:2561–2574.
329. Acevedo ML, Kraus WL. Transcriptional activation by nuclear receptors. *Essays Biochem* 2004;40:73–88.
330. Gill G. Regulation of the initiation of eukaryotic transcription. *Essays Biochem* 2001;37:33–43.
331. Torchia J, Glass C, Rosenfeld MG. Co-activators and co-repressors in the integration of transcriptional responses. *Curr Opin Cell Biol* 1998; 10:373–383.
332. Spiegelman BM, Heinrich R. Biological control through regulated transcriptional coactivators. *Cell* 2004;119:157–167.
333. Tapscott SJ, Davis RL, Thayer MJ, Cheng P-F, Weintraub H, Lassar AB. MyoD1: a nuclear phosphoprotein requiring a myc homology region to convert fibroblasts to myoblasts. *Science* 1988;242:405–411.
334. McKinsey TA, Zhang CL, Olson EN. Signaling chromatin to make muscle. *Curr Opin Cell Biol* 2002;14:763–772.
335. Kumar MS, Owens GK. Combinatorial control of smooth muscle-specific gene expression. *Arterioscler Thromb Vasc Biol* 2003;23: 737–747.
336. Parker MH, Seale P, Rudnicki MA. Looking back to the embryo: defining transcriptional networks in adult myogenesis. *Nat Rev Genet* 2003;4:497–507.
337. Cao D, Wang Z, Zhang CL, Oh J, Xing W, Li S, Richardson JA, Wang DZ, Olson EN. Modulation of smooth muscle gene expression by association of histone acetyltransferases and deacetylases with myocardin. *Mol Cell Biol* 2005;25:364–376.
338. Berk AJ, Schmidt MC. How do transcription factors work? *Genes Dev* 1992;4:151–155.
339. Bluthgen N, Kielbasa SM, Herzel H. Inferring combinatorial regulation of transcription in silico. *Nucleic Acids Res* 2005;33:272–279.
340. Remenyi A, Scholer HR, Wilmanns M. Combinatorial control of gene expression. *Nat Struct Mol Biol* 2004;11:812–815.
341. Kato M, Hata N, Banerjee N, Futcher B, Zhang MQ. Identifying combinatorial regulation of transcription factors and binding motifs. *Genome Biol* 2004;5:R56.
342. Jonas BA, Privalsky ML. SMRT and N-CoR corepressors are regulated by distinct kinase signaling pathways. *J Biol Chem* 2004;279: 54676–54686.
343. Cowley SM, Kang RS, Frangioni JV, Yada JJ, DeGrand AM, Radhakrishnan I, Eisenman RN. Functional analysis of the Mad1-mSin3A repressor-corepressor interaction reveals determinants of specificity, affinity, and transcriptional response. *Mol Cell Biol* 2004; 24:2698–2709.
344. Song H, Hasson P, Paroush Z, Courey AJ. Groucho oligomerization is required for repression in vivo. *Mol Cell Biol* 2004;24:4341–4350.
345. Barolo S, Stone T, Bang AG, Posakony JW. Default repression and Notch signaling: Hairless acts as an adaptor to recruit the corepressors Groucho and dCtBP to Suppressor of Hairless. *Genes Dev* 2002;16:1964–1976.
346. Durst KL, Hiebert SW. Role of RUNX family members in transcriptional repression and gene silencing. *Oncogene* 2004;23:4220–4224.
347. Samuels M, Fire A, Sharp PA. Separation and characterization of factors mediating accurate transcription by RNA polymerase II. *J Biol Chem* 1982;257:14419–14427.
348. Dignam JD, Lebovitz RM, Roeder RG. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* 1983;11:1475–1489.
349. Dynan WS, Tjian R. Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. *Cell* 1983;32:669–680.
350. Gorman CM, Moffat LF, Howard BH. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol Cell Biol* 1982;2:1044–1051.
351. Lim K, Chae C-B. A simple assay for DNA transfection by incubation of the cells in culture dishes with substrates for beta-galactosidase. *Biotechniques* 1989;7:576–579.
352. deWet JR, Wood KV, DeLuca M, Helinski DR, Subramani S. Firefly luciferase gene: structure and expression in mammalian cells. *Mol Cell Biol* 1987;7:725–737.
353. Arnone MI, Dmochowski IJ, Gache C. Using reporter genes to study cis-regulatory elements. *Methods Cell Biol* 2004;74:621–652.
354. Walsh JH, Grossman MI. Gastrin. *N Engl J Med* 1975;292(pt 2): 1377–1384.
355. Liddle RA, Carter JD, McDonald AR. Dietary regulation of rat intestinal cholecystokinin gene expression. *J Clin Invest* 1988;81:2015–2019.
356. Yamada T. Local regulatory actions of gastrointestinal peptides. In: Johnson LR, eds. *Physiology of the gastrointestinal tract.* New York: Raven Press, 1987;131–142.
357. Brand SJ, Stone D. Reciprocal regulation of antral gastrin and somatostatin gene expression by omeprazole-induced achlorhydria. *J Clin Invest* 1988;82:1059–1066.
358. Hakanson R, Axelson J, Ekman R, Sundler F. Hypergastrinaemia evoked by omeprazole stimulates growth of gastric mucosa but not

of pancreas or intestines in hamster, guinea pig and chicken. *Regul Pept* 1988;23:105–115.
359. Wu SV, Sumii K, Tari A, Sumii M, Walsh JH. Regulation of rat antral gastrin and somatostatin gene expression during starvation and after refeeding. *Gastroenterology* 1991;101:1552–1558.
360. Walsh JH, Wu SV. Transcriptional regulation of gastrin and somatostatin synthesis in rats. In: Hakanson R, Sundler F, eds. *Fernstrom foundation series: the stomach as an endocrine organ.* Amsterdam: Elsevier, 1991;223–231.
361. Zavros Y, Rathinavelu S, Kao JY, Todisco A, DelValle J, Weinstock JV, Low MJ, Merchant JL. Treatment of Helicobacter gastritis with interleukin-4 requires somatostatin. *Proc Natl Acad Sci* 2003;100: 12944–12949.
362. Ristevski S. Making better transgenic models: conditional, temporal, and spatial approaches. *Mol Biotechnol* 2005;29:153–164.
363. Jurjus AR, Khoury NN, Reimund JM. Animal models of inflammatory bowel disease. *J Pharmacol Toxicol Methods* 2004;50:81–92.
364. Takagi H, Jhappan C, Sharp R, Merlino G. Hypertrophic gastropathy resembling Menetrier's disease in transgenic mice overexpressing transforming growth factor alpha in the stomach. *J Clin Invest* 1992; 90:1161–1167.
365. Dempsey PJ, Goldenring JR, Soroka CJ, Modlin IM, McClure RW, Lind CD, Ahlquist DA, Pittelkow MR, Lee DC, Sandgren EP, Page DL, Coffey RJ. Possible role of transforming growth factor alpha in the pathogenesis of Ménétrier's disease: supportive evidence from humans and transgenic mice. *Gastroenterology* 1992;103:1950–1963.
366. Donehower LA, Harvey M, Slagle BL, McArthur MJ, Montgomery CA Jr, Butel JS, Bradley A. Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* 1992; 356:215–221.
367. Crooke ST. Antisense strategies. *Curr Mol Med* 2004;4:465–487.
368. Misra RP, Duncan SA. Gene targeting in the mouse: advances in introduction of transgenes into the genome by homologous recombination. *Endocrine* 2002;19:229–238.
369. Seidman MM. Oligonucleotide mediated gene targeting in mammalian cells. *Curr Pharm Biotechnol* 2004;5:421–430.
370. Paroo Z, Corey DR. Challenges for RNAi in vivo. *Trends Biotechnol* 2004;22:390–394.
371. Vanhecke D, Janitz M. Functional genomics using high-throughput RNA interference. *Drug Discov Today* 2005;10:205–212.
372. MacLaren A, Black EJ, Clark W, Gillespie DA. c-Jun-deficient cells undergo premature senescence as a result of spontaneous DNA damage accumulation. *Mol Cell Biol* 2004;24:9006–9018.
373. Feederle R, Delecluse HJ, Rouault JP, Schepers A, Hammerschmidt W. Efficient somatic gene targeting in the lymphoid human cell line DG75. *Gene* 2004;343:91–97.
374. Revzin A. Gel electrophoresis assays for DNA-protein interactions. *Biotechniques* 1989;7:346–355.
375. Mueller PR, Salser SJ, Wold B. Constitutive and metal-inducible protein: DNA interactions at the mouse metallothionein I promoter examined by in vivo and in vitro footprinting. *Genes Dev* 1988;2:412–427.
376. Landolfi NF, Yin X-M, Capra JD, Tucker PW. Protection analysis (or "footprinting") of specific protein-DNA complexes in crude nuclear extracts using methidiumpropyl-EDTA-iron (II). *Biotechniques* 1989; 7:500–504.
377. Garner MM, Revzin A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res* 1981;9:3047–3060.
378. Fried M, Crothers DM. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res* 1981;9:6505–6525.
379. Williams M, Brys A, Weiner AM, Maizels N. A rapid method for determining the molecular weight of a protein bound to nucleic acid in a mobility shift assay. *Nucleic Acids Res* 1992;20:4935–4936.
380. Herr W. Diethyl pyrocarbonate: a chemical probe for secondary structure in negatively supercoiled DNA. *Proc Natl Acad Sci U S A* 1985; 82:8009–8013.
381. Billon N, Carlisi D, Datto MB, van Grunsven LA, Watt A, Wang XF, Rudkin BB. Cooperation of Sp1 and p300 in the induction of the CDK inhibitor p21WAF1/CIP1 during NGF-mediated neuronal differentiation. *Oncogene* 1999;18:2872–2882.
382. Silva CM, Tully DB, Petch LA, Jewell M, Cidlowski JA. Application of a protein-blotting procedure to the study of human glucocorticoid receptor interactions with DNA. *Proc Natl Acad Sci U S A* 1987; 84:1744–1748.
383. Miskimins WK, Roberts MP, McClelland A, Ruddle FH. Use of a protein-blotting procedure and a specific DNA probe to identify nuclear proteins that recognize the promoter region of the transferrin receptor gene. *Proc Natl Acad Sci U S A* 1985;82:6741–6744.
384. Singh H, Clerc RG, LeBowitz JH. Molecular cloning of sequence-specific DNA binding proteins using recognition site probes. *Biotechniques* 1989;7:252–261.
385. Das PM, Ramachandran K, vanWert J, Singal R. Chromatin immunoprecipitation assay. *Biotechniques* 2004;37:961–969.
386. Bernstein BE, Humphrey EL, Liu CL, Schreiber SL. The use of chromatin immunoprecipitation assays in genome-wide analyses of histone modifications. *Methods Enzymol* 2004;376:349–360.
387. Weinmann AS, Farnham PJ. Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods* 2002;26:37–47.
388. Meister G, Tuschl T. Mechanisms of gene silencing by double-stranded RNA. *Nature* 2004;431:343–349.
389. Mockler TC, Ecker JR. Applications of DNA tiling arrays for whole-genome analysis. *Genomics* 2005;85:1–15.
390. Mantripragada KK, Buckley PG, de Stahl TD, Dumanski JP. Genomic microarrays in the spotlight. *Trends Genet* 2004;20:87–94.
391. Simon R, Mirlacher M, Sauter G. Tissue microarrays. *Biotechniques* 2004;36:98–105.
392. Cunliffe VT. Memory by modification: the influence of chromatin structure on gene expression during vertebrate development. *Gene* 2003;305:141–150.
393. Buchholz M, Gress TM. Application of DNA array analyses in the management of gastrointestinal cancer patients. *Dig Dis* 2003;21: 309–314.
394. Cowell JK. High throughput determination of gains and losses of genetic material using high resolution BAC arrays and comparative genomic hybridization. *Comb Chem High Throughput Screen* 2004; 7:587–596.
395. Huang RP. Protein arrays, an excellent tool in biomedical research. *Front Biosci* 2003;8:d559–d576.
396. Ng JH, Ilag LL. Biomedical applications of protein chips. *J Cell Mol Med* 2002;6:329–340.
397. Brink C. Promoter elements in endocrine pancreas development and hormone regulation. *Cell Mol Life Sci* 2003;60:1033–1048.
398. Cha-Molstad H, Keller DM, Yochum GS, Impey S, Goodman RH. Cell-type-specific binding of the transcription factor CREB to the cAMP-response element. *Proc Natl Acad Sci U S A* 2004;101:13572–13577.
399. Burbach JP. Regulation of gene promoters of hypothalamic peptides. *Front Neuroendocrinol* 2002;23:342–369.
400. Hahm SH, Eiden LE. Cis-regulatory elements controlling basal and inducible VIP gene transcription. *Ann N Y Acad Sci* 1998;865:10–26.
401. Merchant JL, Tucker TP, Zavros Y. Inducible regulation of gastrin gene expression. In: Merchant JL, Buchan AM, Wang TC, eds. *Gastrin in the new millennium.* Los Angeles: CURE Foundation, 2004;55–69.
402. Mutoh H, Ratineau C, Ray S, Leiter AB. Review article: transcriptional events controlling the terminal differentiation of intestinal endocrine cells. *Aliment Pharmacol Ther* 2000;14(suppl 1):170–175.
403. Rindi G, Ratineau C, Ronco A, Candusso ME, Tsai M, Leiter AB. Targeted ablation of secretin-producing cells in transgenic mice reveals a common differentiation pathway with multiple enteroendocrine cell lineages in the small intestine. *Development* 1999;126:4149–4156.
404. Lay JM, Bane G, Brunkan CS, Davis J, Lopez-Diaz L, Samuelson LC. Enteroendocrine cell expression of a cholecystokinin gene construct in transgenic mice and cultured cells. *Am J Physiol Gastrointest Liver Physiol* 2005;288:G354–G361.
405. Zorio DA, Bentley DL. The link between mRNA processing and transcription: communication works both ways. *Exp Cell Res* 2004; 296:91–97.
406. Proudfoot N. Ending the message is not so simple. *Cell* 1996;87: 779–781.
407. Bernstein P, Peltz SW, Ross J. The Poly(A)-Poly(A)-binding protein complex is a major determinant of mRNA stability in vitro. *Mol Cell Biol* 1989;9:659–670.
408. Vassalli J-D, Huarte J, Belin D, Gubler P, Vassalli A, O'Connell ML, Parton LA, Rickles RJ, Strickland S. Regulated polyadenylation controls mRNA translation during meiotic maturation of mouse oocytes. *Genes Dev* 1989;3:2163–2171.
409. Gallie DR. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev* 1991;5:2108–2116.
410. Zarudnaya MI, Kolomiets IM, Potyahaylo AL, Hovorun DM. Downstream elements of mammalian pre-mRNA polyadenylation

signals: primary, secondary and higher-order structures. *Nucleic Acids Res* 2003;31:1375–1386.
411. Takagaki Y, Ryner LC, Manley JL. Four factors are required for 3′-end cleavage of pre-mRNAs. *Genes Dev* 1989;3:1711–1724.
412. Gilmartin GM, Nevins JR. An ordered pathway of assembly of components required for polyadenylation site recognition and processing. *Genes Dev* 1989;3:2180–2189.
413. Wahle E. A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* 1991; 66:759–768.
414. Kirsh AL, Groudine M, Challoner PB. Polyadenylation and U7 snRNP-mediated cleavage: alternative modes of RNA 3′ processing in two avian histone H1 genes. *Genes Dev* 1993;3:2172–2179.
415. Darnell J, Lodish H, Baltimore D. *Molecular cell biology*. New York: Scientific American Books, 1990.
416. Calvo O, Manley JL. Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev* 2003;17:1321–1327.
417. Krainer AR, Maniatis T. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 necessary for pre-mRNA splicing in vitro. *Cell* 1985;42:725–736.
418. Krämer A, Frick M, Keller W. Separation of multiple components of HeLa cell nuclear extracts required for pre-messenger RNA splicing. *J Biol Chem* 1987;262:17630–17640.
419. Maniatis T, Reed R. The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature* 1987;325:673–678.
420. Nilsen TW. The spliceosome: the most complex macromolecular machine in the cell? *Bioessays* 2003;25:1147–1149.
421. Berget SM. Exon recognition in vertebrate splicing. *J Biol Chem* 1995; 270:2411–2414.
422. Reed R, Maniatis T. A role for exon sequences and splice-site proximity in splice-site selection. *Cell* 1986;46:681–690.
423. Nelson KK, Green MR. Splice site selection and ribonucleoprotein complex assembly during in vitro pre-mRNA splicing. *Genes Dev* 1988;2:319–329.
424. Jacob M, Gallinaro H. The 5′ splice site: phylogenetic evolution and variable geometry of association with U1RNA. *Nucleic Acids Res* 1989; 17:2159–2180.
425. Krämer A. Presplicing complex formation requires two proteins and U2 snRNP. *Genes Dev* 1988;2:1155–1167.
426. Nelson KK, Green MR. Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes Dev* 1989;3:1562–1571.
427. Fabrizio P, McPheeters DS, Abelson J. In vitro assembly of yeast U6 snRNP: a functional assay. *Genes Dev* 1989;3:2137–2150.
428. Reed R. The organization of 3′ splice-site sequences in mammalian introns. *Genes Dev* 1989;3:2113–2123.
429. Madhani HD, Guthrie C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* 1992;71:803–817.
430. Weiner AM. mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism? *Cell* 1993;72:161–164.
431. Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G. Multiple links between transcription and splicing. *Rna* 2004;10: 1489–1498.
432. Shin C, Manley JL. Cell signalling and the control of pre-mRNA splicing. *Nat Rev Mol Cell Biol* 2004;5:727–738.
433. Darnell JE Jr. The processing of RNA. *Sci Am* 1983;249:90–100.
434. Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 2003;72:291–336.
435. Kashiwabara S, Zhuang T, Yamagata K, Noguchi J, Fukamizu A, Baba T. Identification of a novel isoform of poly(A) polymerase, TPAP, specifically present in the cytoplasm of spermatogenic cells. *Dev Biol* 2000;228:106–115.
436. Hoeffler JP, Meyer TE, Waeber G, Habener JF. Multiple adenosine 3′,5′-monophosphate response element DNA-binding proteins generated by gene diversification and alternative exon splicing. *Mol Endocrinol* 1990;4:920–930.
437. Helfman DM, Roscigno RF, Mulligan GJ, Finn LA, Weber KS. Identification of two distinct intron elements involved in alternative splicing of beta-tropomyosin pre-mRNA. *Genes Dev* 1990; 4:98–110.
438. Breitbart RE, Andreadis A, Nadal-Ginard B. Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu Rev Biochem* 1987;56:467–495.
439. Klausner RD, Harford JB. Cis-Trans models for post-transcriptional gene regulation. *Science* 1989;246:870–872.
440. Templeton DM, Liu Y. Genetic regulation of cell function in response to iron overload or chelation. *Biochim Biophys Acta* 2003;1619: 113–124.
441. Hentze MW, Caughman SW, Rouault TA, Barriocanal JG, Dancis A, Harford JB, Klausner RD. Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science* 1987;238:1570–1573.
442. Casey JL, Hentze MW, Koeller DM, Caughman SW, Rouault TA, Klausner RD, Harford JB. Iron-responsive elements: regulatory RNA sequences that control mRNA levels and translation. *Science* 1988; 240:924–928.
443. Pemberton LF, Paschal BM. Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* 2005;6:187–198.
444. Hoelz A, Blobel G. Cell biology: popping out of the nucleus. *Nature* 2004;432:815–816.
445. Lusk CP, Makhnevych T, Wozniak RW. New ways to skin a kap: mechanisms for controlling nuclear transport. *Biochem Cell Biol* 2004;82:618–625.
446. Unwin PNT, Milligan RA. A large particle associated with the perimeter of the nuclear pore complex. *J Cell Biol* 1982;93:63–75.
447. Hinshaw JE, Carragher BO, Milligan RA. Architecture and design of the nuclear pore complex. *Cell* 1992;69:1133–1141.
448. Rout MP, Aitchison JD, Magnasco MO, Chait BT. Virtual gating and nuclear transport: the hole picture. Trends Cell Biol 2003;13: 622–628.

CHAPTER 2

# Translation and Posttranslational Processing of Gastrointestinal Peptides

Cheryl E. Gariepy and Chris J. Dickinson



## TRANSLATION

Translation is the complex process by which a sequence of codons of messenger ribonucleic acid (mRNA) directs the synthesis of a polypeptide chain. Beyond the sequence of codons, the mRNA contains untranslated regions (UTRs) with structural and regulatory sequences that determine its translational fate. Translation involves hundreds of molecules including mRNA, transfer RNA (tRNA), ribosomal RNA, activation enzymes, and many RNA-binding proteins, as well as energy in the form of guanosine triphosphate (GTP) and adenosine triphosphate (ATP).

To be translated into protein, mRNA must contain, in addition to a string of codons, information that specifies nuclear export, translation, and stability. Much of this information is communicated by specific RNA-binding proteins. These proteins first associate with pre-mRNA (primary transcripts of genomic DNA-containing exons and introns) cotranscriptionally and undergo a dynamic series of rearrangements involving the binding and dissociation of numerous proteins throughout the life of mRNA. The mRNA nucleoprotein complex (mRNP) communicates information to the cytoplasm about the structure of the gene from which the mRNA was formed and the processing steps experienced by the mRNA. The mRNP therefore carries significantly greater information than the sequence of the mRNA itself (1).

C. E. Gariepy and C. J. Dickinson: Division of Pediatric Gastroenterology, University of Michigan, Ann Arbor, Michigan 48109.

*Physiology of the Gastrointestinal Tract, Fourth Edition*, edited by Leonard R. Johnson. Academic Press, 2006.

### Initiation

Translation requires the positioning of an elongation-competent 80S ribosome at the initiation codon (AUG). Binding of the small (40S) ribosomal subunit at the 5′ end of the mRNA is rate limiting and requires energy in the form of ATP. It then forms a 43S preinitiation complex with eukaryotic initiation factors (eIFs) 3, 1, 1A, and 5, and a ternary complex including the methionine-loaded initiator tRNA that will recognize the AUG codon and eIF2 that is coupled to GTP. The preinitiation complex recognizes the mRNA by the binding of eIF3 to the eIF4 protein complex associated with the 5′ cap structure (eIF4F) (2). The eIF4F protein complex contains an enzyme (eIF4A) that unwinds RNA duplexes, allowing the 43S complex to bind and scan the mRNA, and a scaffold protein (eIF4G) that serves as a platform for the assembly of other proteins and interacts with the poly(A)-binding protein. This interaction is thought to loop the mRNA and bring the 3′ UTR in close proximity to the 5′ end of the mRNA (3). This provides a means by which sequences in the 3′ UTR can regulate translation initiation. Most known translational regulatory sequences are found within the 3′ UTR.

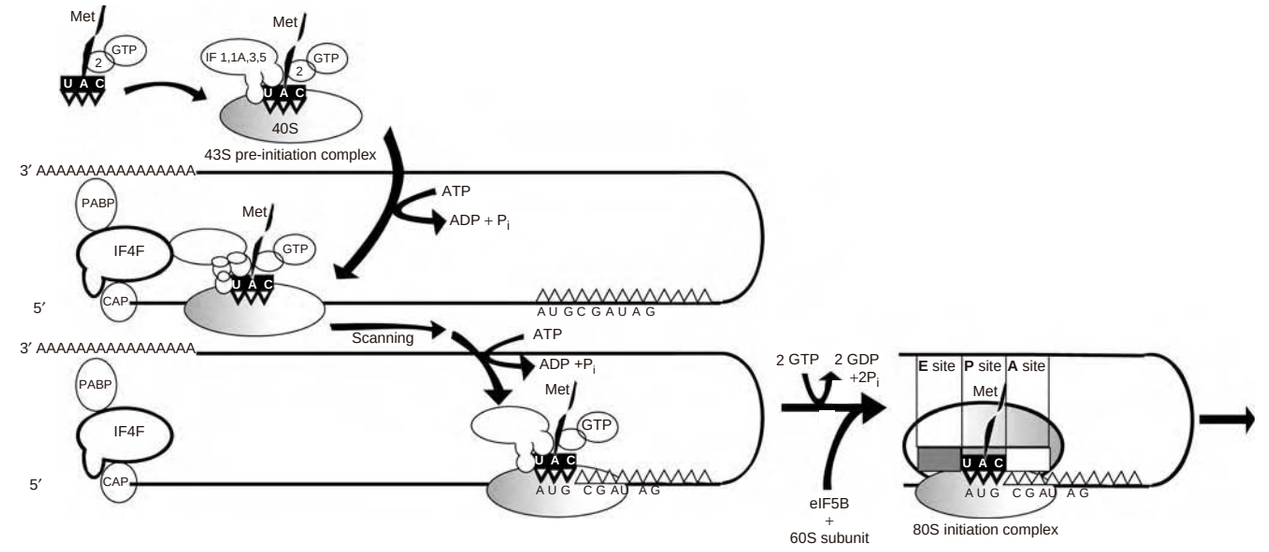
The 43S complex recognizes the initiation codon through the formation of base pairs (bp) between the initiator tRNA and the start codon. Subsequently, eIF2-bound GTP undergoes hydrolysis, a reaction that is necessary for the 60S ribosomal subunit to join the initiation complex. This appears to release most of the initiation factors from the small ribosomal subunit, leaving the initiator tRNA associated with the ribosome (in the P site). Formation of the 80S initiation complex capable of catalyzing the formation of a peptide bond occurs with the hydrolysis of a second molecule of GTP on eIF5B (Fig. 2-1A).

### Regulation of Initiation

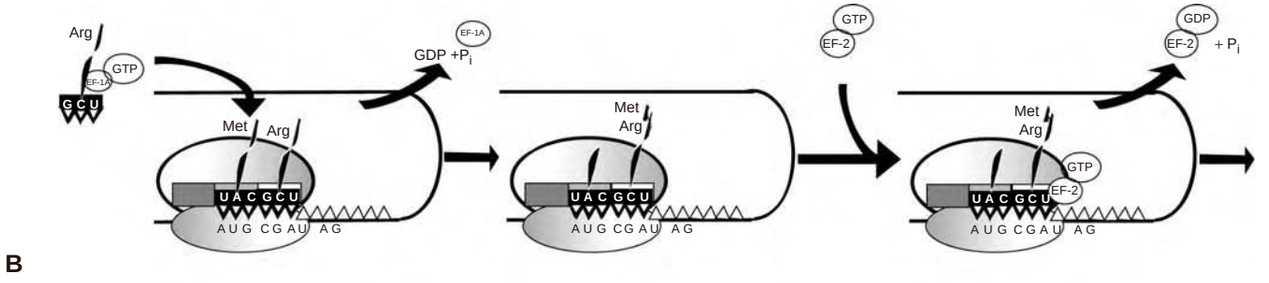
Although the specific translational regulatory mechanisms active in peptide hormone synthesis are not yet clear, translation is generally controlled at the initiation step where regulation may be global or mRNA specific. Global control of mRNA translation generally occurs through changes in the phosphorylation state of initiation factors or regulators that interact with them. Proteolytic cleavage of translation factors can also reduce translation of all mRNA species within the cell. mRNA-specific regulation of translation can be achieved by steric blockage, interference with the eIF4F complex, and cap-independent inhibition of the early initiation steps. Steric blockage refers to the binding of regulatory proteins to message-specific response elements that results in insufficient space for the binding of critical initiation complex proteins (4–7). Interference with the eIF4F complex is achieved by mRNA-specific binding proteins that block eIF4E recognition by eIF4G (8–11). Cap-independent inhibition of translation refers to proteins that bind to specific sites in both the 5′ and 3′ UTRs and recruit corepressors to the 3′ UTR. This affects stable association of the small ribosomal subunit with the mRNA (12–14).

Translation can also be controlled later in the initiation process. RNA-binding proteins have been described that prevent the binding of the 60S ribosomal subunit to the 40S subunit at the initiation codon, apparently through interference with initiation factors (15). The existence of more than one open reading frame on an mRNA and the sequence distance between the open reading frames can also play a significant role in determining the likelihood of translation. For example, amino acid deprivation reduces global protein

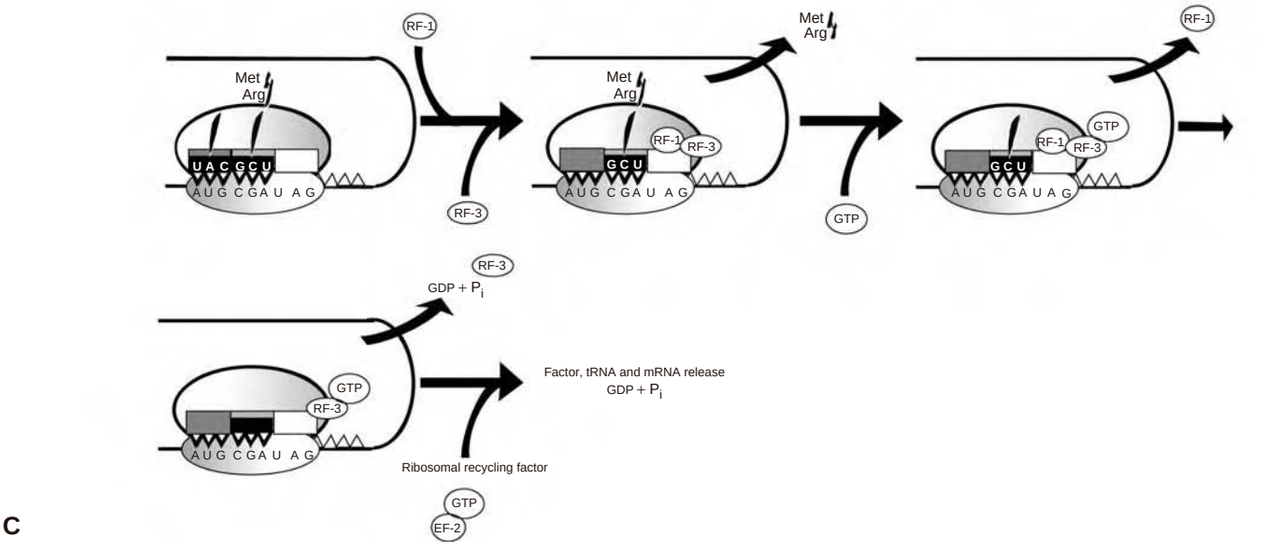
**FIG. 2-1.** Translation of messenger RNA (mRNA) into protein, highly simplified. Some translation-initiation factors are omitted (see Preiss and Hentze [526], Ramakrishnan [527], and Hershey [528] for more complete descriptions). **(A)** Cap-mediated initiation: The methionine-containing ternary complex (methionine-loaded transfer RNA [tRNA], eIF2, and GTP) binds to the 40S ribosomal subunit and other initiation factors (eukaryotic initiation factor-1 [eIF1], 1A, 3, and 5) to form the 43S preinitiation complex. The preinitiation complex recognizes the mRNA through the binding of eIF3 to eIF4 in the cap-binding complex. The cap-binding complex contains eIF4A, an RNA helicase that unwinds the secondary structure of the mRNA during the subsequent scanning step. The cap-binding complex also contains eIF4G, which contacts the poly(A)-binding protein (PABP). This contact is thought to bring the 3′ region of the mRNA in close proximity to the 5′ cap. The 43S preinitiation complex scans the mRNA from 5′ to 3′ until the initiation codon, AUG, is encountered. Stable binding of the preinitiation complex to the AUG codon yields the initiation complex. Subsequent joining of the 60S ribosomal subunit results in the formation of the 80S initiation complex. AUG recognition and the joining of the 60S ribosomal subunit both trigger GTP hydrolysis. The 80S complex contains an aminoacylated initiator tRNA in the P site of the ribosome and an empty A site. It now is competent to catalyze the formation of the first peptide bond. **(B)** Elongation: A ternary complex containing aminoacylated tRNA and the correct anticodon is brought into the A site of the ribosome. Codon–anticodon recognition leads to guanosine triphosphate (GTP) hydrolysis. This allows for conformational changes within the tRNA and the ribosome. Peptide bond formation (deacylation of the P site tRNA and the transfer of the peptide chain to the A site tRNA) then occurs. Translocation of the tRNA and the mRNA is facilitated by a GTPase, eukaryotic elongation factor-2 (eEF2). The ribosome is then ready for the next round of elongation, with a deacylated tRNA in the E site, peptidyl tRNA in the P site, and an empty A site. **(C)** Termination: When a stop codon on the mRNA is encountered in the A site, eukaryotic release factor-1 (eRF1) binds to the ribosome A site and triggers the release of the peptide chain from the tRNA in the P site. eRF3 then binds GTP and promotes dissociation of eRF1 from the ribosome. Hydrolysis of GTP is required for subsequent release of eRF3. The ribosome is then left with mRNA and a deacylated tRNA in the P site. The ribosomal releasing factor, together with eEF2 and GTP, is required to disassemble the complex and prepare the ribosome for a new round of protein synthesis. Much of the mechanism of mRNA, translation factor, and subunit release after peptide chain termination remain to be determined. GDP, guanosine diphosphate.


Met
GTP
2
U A C
IF 1,1A,3,5
40S
43S pre-initiation complex
3′ AAAAAAAAAAAAAAAA
PABP
IF4F
CAP
5′
ATP
ADP + Pi
A U G C G A U A G
Scanning
ADP +Pi
2 GTP
2 GDP
+2Pi
E site
P site
A site
eIF5B
+
60S subunit
80S initiation complex

A


Arg
GTP
EF-1A
G C U
GDP +Pi
Met
U A C G C U
A U G C G A U A G
EF-2
GDP
+ Pi

B


RF-1
RF-3
GTP
GDP + Pi
Factor, tRNA and mRNA release
GDP + Pi
Ribosomal recycling factor
EF-2

C

synthesis by phosphorylation of eIF2a, which blocks GDP-GTP exchange and reconstitution of the functional ternary complex. Paradoxically, the same modification increases the translation of some mRNA that have upstream open reading frames. It appears that the 60S ribosomal subunit dissociates at the stop codon of the first open reading frame and the 40S subunits remain associated with the mRNA and resumes scanning. The 40S subunit must acquire an active ternary complex during scanning to translate downstream open reading frames. The probability of translating the most 3′ open reading frame therefore depends on the distance (scanning time) between the open reading frames and the availability of amino acids within the cell (16).

Internal ribosome entry sites (IRESs) mediate translation initiation independent of the cap structure by recruiting the ribosome directly to an internal position of the mRNA (17). Both structural features and short-sequence elements appear to be involved in ribosome recruitment in eukaryotic IRESs. Exactly how these motifs combine to promote internal initiation remains to be determined. The IRES appears to be a complex RNA scaffold that contains multiple sites for interaction with components of the translational apparatus. Structural domains have been identified that interact with the initiation factors eIF4G and eIF4B (18,19), with eIF3 (20,21), or directly with the 40S ribosome subunit at multiple sites. An IRES is also described that can assemble an 80S ribosome at its initiation codon without the aid of any initiation factors or an initiator tRNA (22). A growing body of evidence exists to support the hypothesis that cellular IRESs are involved in the regulation of gene expression under physiologic conditions during which the efficiency of cap-dependent protein synthesis is greatly reduced. IRESs enable cells to respond to these conditions against the background of a general reduction in protein synthesis.

### Elongation

Each ribosomal subunit has three binding sites for tRNA: designated the A (aminoacyl) site, which accepts the incoming aminoacylated tRNA; P (peptidyl) site, which holds the tRNA with the nascent peptide chain; and E (exit) site, which holds the deacylated tRNA before it leaves the ribosome. The end of the initiation process leaves an aminoacylated initiator tRNA in the P site of the ribosome and an empty A site, which serves to start the elongation process. Aminoacylated tRNA is brought into the A site as a ternary complex with eukaryotic elongation factor-1A (eEF1A) and GTP. Correct codon-anticodon interactions result in conformational changes in the ribosome that stabilize tRNA binding and trigger GTP hydrolysis by eEF1A. This leads to the release of the aminoacyl end of the A site tRNA by eEF1A; the tRNA then swings into the peptidyl transferase site of the large subunit in a process called accommodation. The peptide bond is formed through deacylation of the P site tRNA and the transfer of the peptide chain to the A site tRNA. The ribosome then has a deacylated tRNA in the P site and peptidyl tRNA in the A site. Translocation of tRNA and mRNA is facilitated by eEF2, which is also a GTPase. The ribosome is then ready for the next round of elongation, with deacylated tRNA in the E site, peptidyl tRNA in the P site, and an empty A site ready to receive the next cognate ternary complex (see Fig. 2-1B).

### Termination

Termination begins when a stop codon (UAA, UGA, or UAG) is encountered in the A site mRNA. Stop codons are recognized by eukaryotic release factor-1 (eRF1). The GTPase eRF3 then binds the complex of eRF1 bound to the ribosome. Binding of eRF1 to the ribosome at the stop codon A site triggers the hydrolysis and release of the peptide chain from the tRNA in the P site. Hydrolysis of peptidyl tRNA by eRF1 is required for binding of GTP to eRF3 on the ribosome. This, in turn, leads to a conformational change in eRF3 that has high affinity for ribosomes and the dissociation of eRF1 from the ribosome. Hydrolysis of GTP is required for subsequent dissociation of eRF3 from the ribosome (see Fig. 2-1C) (23,24).

### Localized Translation Regulation

In addition to regulation of the initiation process, mRNA-specific translation regulation also occurs regionally in polarized cells. This is clearly demonstrated in neural tissues where stimulation of synapses induce the polyadenylation and translation of cytoplasmic polyadenylation element-containing, but not cytoplasmic polyadenylation element-lacking, mRNA stored in dendrites (25). This allows the generation of protein gradients emanating from particular positions in cells or the restriction of protein expression to a specific region and is a potential mechanism by which a cell may modulate its response to repeated, directional stimuli.

### RNA Silencing

Small RNA molecules regulate mRNA-specific translation either by translational repression, in the case of microRNA (miRNA) (26–28), or by mediating the degradation of the target mRNA, in the case of small interfering RNA (siRNA) (29,30). The functional difference between miRNA and siRNA (both about ~22 nucleotides in length) depends on the degree of complementation between the small RNA molecule and the mRNA target (31,32). miRNA hybridize by incomplete base paring, usually to several sites in the 3′ UTR of target mRNA. siRNA show perfect complementation to the target mRNA. miRNA and siRNA have distinctly different origins within the nucleus, but they have common RNA-binding proteins (33). It is unclear whether a single type of small RNA-protein complex can mediate both

target mRNA cleavage and translational inhibition (see review by Sontheimer [34]).

### Other Regulators of Messenger RNA Stability

Regulation of the rate of decay is an important control point in determining the abundance of an mRNA species, and decay rates of individual mRNA differ widely and can be differentially affected by environmental cues. Several sequence elements can regulate the rate of turnover of a transcript by attracting specific binding proteins that can either destabilize or stabilize the transcript. The strength of the association of these binding proteins can be modified by changes in the cellular environment. The principal mRNA-degradation pathway begins with removal of the 3′ poly(A) tail. Interaction of the cap proteins and the poly(A)-binding proteins with the translational machinery likely protects the 5′ and the 3′ end of the mRNA from attack by deadenylases and decapping enzymes (35,36). This means that translation and mRNA decay are linked. Support for this comes from studies demonstrating that inhibition of translation initiation destabilized mRNA (37) and inhibition of translation elongation (with cycloheximide) promotes mRNA stability (38). The nonsense-mediated decay pathway further links translation to mRNA turnover. This pathway ensures that mRNA with premature stop codons are not translated. To be recognized as premature, a termination codon must lie upstream of the last intron (39–43). Exon-exon junctional complex proteins, which mark the position of exon-exon junctions in the mature mRNA, may play an important role in surveillance for potentially deleterious nonsense mutations (44).

## POSTTRANSLATIONAL PROCESSING

Although it would appear that the translation of polypeptide hormones is similar to that of other eukaryotic proteins, the posttranslational processing of prohormones is unique. Since the initial discovery of proinsulin (45), it has been evident that the synthesis of polypeptide hormones of the gut involves a series of modification steps after the initial translation of the gene product that are distinct from the biosynthesis of other cellular proteins. These modifications, achieved via a variety of posttranslational processing reactions, may enlarge or diminish the size of the peptide precursor, but, in general, they result in the formation of biologically active and physiologically relevant products. Efforts to determine the nature and mechanisms of peptide hormone posttranslational processing reactions were greatly facilitated by the development of molecular biological techniques that permitted the deduction of peptide precursor sequences. Information on precursor structure has led to the development of molecular probes that can be used to characterize individual processing reactions, as well as patterns of processing reactions for groups of related peptides. Application of these probes to ultrastructural studies has provided information on the cellular compartments in which processing reactions take place. *In vitro* reconstitution experiments have led to the elucidation of some of the mechanisms responsible for the transport of peptide precursors between cellular compartments. Development of techniques to isolate and culture functionally intact peptide-secreting cells has permitted physiologists to examine the sequence and dynamics of the complete posttranslational modification and activation process for given peptides. Many of the enzymes responsible for prohormone processing have now been isolated. Coexpression or deletion, or both, of these enzymes within cells has allowed for elucidation of their activities for multiple prohormone substrates.

Previously, it was thought that proteins exited from cells via two distinct pathways: the constitutive or the regulated secretory pathways (46,47). As has been the case with other biological systems, more recent evidence suggests that there might be overlap between these pathways (48). Generally, however, the constitutive pathway is reserved for those secreted proteins that are not stored in the cell and usually do not undergo extensive posttranslational processing, as seen with the products of fibroblasts and hepatocytes. Proteins secreted constitutively exit the cell soon after synthesis on the ribosome. Polypeptide hormones, however, enter the regulated pathway of secretion in most neuroendocrine cells. These cells are capable of storing secretory products for hours or days in electron-dense secretory vesicles and releasing them on stimulation. The intracellular pathways and organelles involved in this pathway were first defined in studies (49) in the exocrine pancreas demonstrating that polypeptides are initially synthesized on the rough endoplasmic reticulum (ER), transported to the Golgi apparatus, and finally placed into secretory granules (Fig. 2-2). On cell stimulation, these secretory granules or vesicles fuse in a calcium-dependent manner with the cell membrane to release their contents into the extracellular milieu. This chapter reviews the enormous progress made in recent years in elucidating the mechanisms for posttranslational processing of gastrointestinal peptide hormones, and then presents a detailed analysis of one hormone, gastrin.

## TRANSPORT INTO THE ENDOPLASMIC RETICULUM

Polypeptide hormones are synthesized ribosomally from the amino-terminal end and enter the secretory pathway via translocation into the ER. This process is of critical importance to both prokaryotes and eukaryotes. Thus, it has been thoroughly examined by several notable scientists including Gunther Blobel, who won a Noble Prize for his work in this area (50). In summary, the first few amino acids of the preprohormone, translated from the leader sequence of the specific mRNA, are called the signal peptide (Fig. 2-3) (51). This peptide (designated as the presequence in preprohormones) is not secreted under normal circumstances but serves as a means of translocating the newly synthesized and

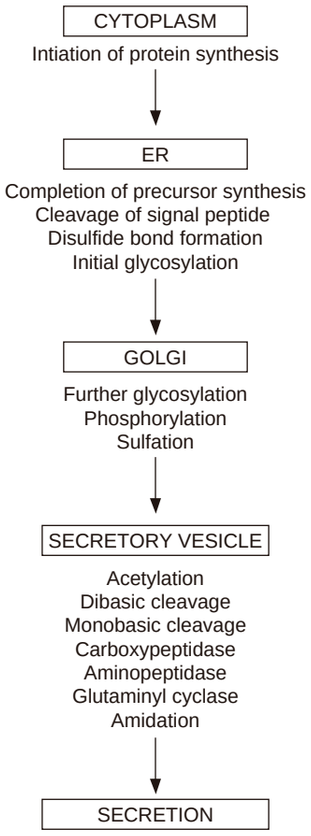



**FIG. 2-2.** Intracellular location of posttranslational processing steps. ER, endoplasmic reticulum.

gradually elongating polypeptide chain into the ER (52–54). After emerging from the ribosome, the signal peptide binds to the signal recognition particle (SRP) in the cytoplasm after chain elongation has produced a preprohormone of approximately 50 to 60 amino acids (55–57). This binding results in an arrest of translation, and the SRP initiates the translocation of the nascent polypeptide by binding to the SRP receptor or docking protein located on the cytosolic side of the ER (58–60). The SRP is then released, and the translocation of the peptide continues through a protein channel across the ER membrane (53,59,61–63). The signal peptide is later cleaved by a specific enzyme (signal peptidase) located on the inner membrane of the ER (64). The individual components of these ER translocation events are described in more detail in the next section.

### Signal Peptides

The signal peptide, or *pre*sequence, constituted by the amino-terminal 20 to 30 amino acids of a newly synthesized polypeptide chain, directs the translocation of the polypeptide into the ER lumen. There seems to be little primary amino acid homology in the signal peptides of the known gastrointestinal hormone precursors. However, three general characteristics are shared: (1) a positively charged amino-terminal region of 1 to 10 amino acids, (2) a central hydrophobic region of 7 to 17 amino acids, and (3) a more polar region that often contains an α helix breaking proline or glycine residue, as well as uncharged residues that determine the cleavage site and complex pattern of amino acids adjacent to the site of cleavage between the signal peptide and the prohormone (65,66). The secondary structure of these peptides can assume several different conformations including α helices and β-pleated sheets, depending on the environment (67). Recently, analysis of the new, extensive protein databases has allowed investigators to accurately predict signal peptides (68). The positively charged amino terminus appears to be important in the release of the SRP once docking of the nascent peptide to the ER has occurred. Mutations in this area that result in a net negative charge interfere with both export and synthesis of secretory proteins in prokaryotes (69), although this does not appear to be the case in eukaryotic systems (70). Mutations that substitute polar or charged amino acids for the amino acids present in the hydrophobic region of the signal peptide result in impaired binding of the nascent peptide chain to the SRP (57). Thus, translation is complete, but export of newly synthesized protein is inhibited (71). Initially, it was thought that conservative substitutions of one hydrophobic amino acid for another (e.g., glycine for valine) did not alter the recognition between the SRP and the signal peptide (72). More recently, others have noted that even small changes in the central hydrophobic core can alter SRP binding (73,74). However, in these cases, translocation across the ER membrane still occurs through an unknown mechanism. SRP binding is not dependent on the presence of a net positive charge at the amino terminus or on any identifiable features at the carboxyl terminus (75,76). There do not appear to be any specific structural requirements for the site of signal peptide cleavage, although the carboxyl-terminal amino acid of the signal peptide usually has a small uncharged side chain such as alanine (65,77) (see Signal Peptidase later in this chapter).

After translocation through the ER membrane pore, the signal peptide can loop back through the membrane. The signal peptide is then cleaved at its c terminus by signal peptidase. However, recent studies have shown that the signal peptide can be cleaved further by a signal peptide peptidase to release the amino terminal fragment into the cytosol (78). In the case of prolactin, this fragment then binds to calmodulin in a calcium-dependent manner (79). Currently, the full biological implications of this finding are unknown. In addition, it is not known if this applies to multiple other peptide hormones.

### Signal Recognition Particle

SRP has three known functions: signal peptide recognition, elongation arrest, and promotion of translocation (55,56, 80–83). This particle consists of six polypeptide components with molecular weights of 72, 68, 54, 19, 14, and 9 kDa,

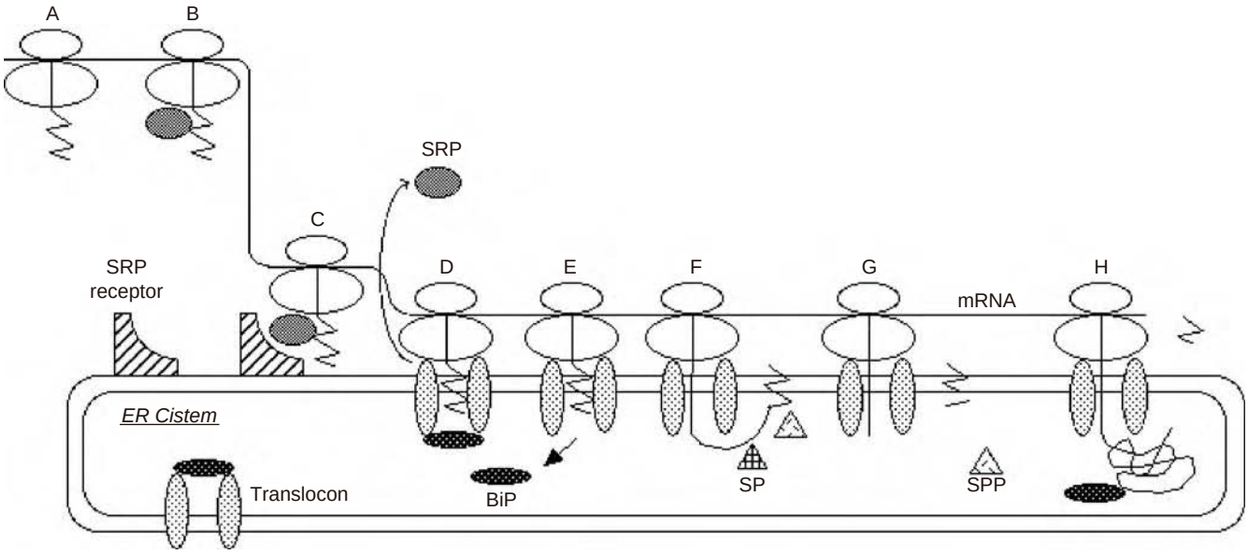



**FIG. 2-3.** Signal peptides. **(A)** Ribosome binds to the messenger RNA (mRNA) and translation begins at the amino terminus. **(B)** Signal sequence emerges from the ribosome and binds to the signal recognition particle (SRP), which induces an arrest of translation. **(C)** SRP-ribosome complex binds to the SRP receptor or docking protein located on the endoplasmic reticulum (ER) membrane. **(D)** The translocon or protein pore binds to the ribosome, releasing the SRP and its receptor. **(E)** With binding of the ribosome to the cytosolic side of the translocon, BiP is released from the lumenal side of the translocon. **(F)** Translation resumes at the carboxyl terminus, and the signal peptide can be reinserted into the ER membrane. **(G)** Signal peptide is cleaved for the prohormone by signal peptidase (SP), and the signal peptide fragment is released into the cytosol via signal peptide peptidase (SPP) cleavage. **(H)** Translation continues until the entire precursor is located within the ER cistern and properly folded, often in association with chaperones such as BiP.

as well as a 7SL RNA (80,84). Each component is held together in a defined tertiary structure by $Mg^{2+}$ ions and is essential for the functions of the SRP. The 19- and 54-kDa proteins exist as monomers, but heterodimers of the 9- and 14-kDa proteins and the 68- and 72-kDa proteins are formed (85). The 54-kDa protein contains a series of amphipathic helices with methionine residues, located predominantly on one face, that appear to be important for the binding to the hydrophobic region of the signal peptide (83,86–88). Interestingly, SRP will not bind to signal peptides that are not tethered to a ribosome, although the peptide region responsible for the SRP–ribosome interaction is not known. It appears that the 54-kDa protein binds to the 7SL RNA through the 19-kDa protein that binds directly to the middle of the RNA strand (88). In addition, the 7SL RNA contains 5′ and 3′ Alu-like elements that bind to each other and the 9/14-kDa protein heterodimer (89–91). The 9/14-kDa protein heterodimer mediates elongation arrest of translation, but plays no role in the translocation process (91,92). The 68/72-kDa heterodimer binds to the middle segment of RNA close to the 19-kDa binding site and appears to mediate the binding of SRP to its receptor (93,94). Thus, the 68/72-kDa heterodimer is not involved in elongation arrest but serves to aid in translocation.

The 54-kDa protein binds GTP in concert with binding to the signal peptide (80,82,95–97). An additional GTP is required on binding of the SRP to the SRP receptor (96). When the SRP/SRP receptor complex associates with the ER membrane or translocon there is a subsequent release of GDP (98). The hydrolysis of GTP releases the SRP from the signal peptide and allows translation to proceed.

### Signal Recognition Particle Receptor

The SRP receptor is located on the cytosolic side of the ER and binds to the SRP–ribosome complex, but not to free SRP as noted earlier. The SRP receptor plays an important role in termination of the elongation arrest and in the translocation of polypeptides into the ER lumen (59). The SRP receptor is a heterodimeric protein consisting of a 30-kDa integral membrane protein (β subunit) and 72-kDa α subunit that possesses domains that are homologous to GTP-binding proteins and the GTP-binding region of the 54-kDa SRP protein (88,95,99–104). The 72-kDa α subunit of the SRP receptor binds to SRP, and GTP is necessary to release SRP from the signal peptide–ribosome complex (95,105).

### Endoplasmic Reticulum Membrane Protein Channel or Translocon

There have been numerous theories about whether the nascent polypeptide chain is transported directly across the

lipid bilayer or in the aqueous environment of a protein channel. Although the initial thought was that the hydrophobic core of the signal peptide would allow for direct translocation across the membrane, there is a large transmembrane channel that is opened by the presence of signal peptides (106–108). The eukaryotic ER translocon is a heterotrimer (Sec61α, Sec62β, and Sec61γ) estimated to have a diameter of about 30 Å (109–111). The size of the pore is too small for folded proteins, thus ensuring that only nascent, unfolded proteins can enter the ER lumen. The lumenal side of the translocon is sealed by a protein, BiP, that aids in protein folding after passage through the membrane (112). On binding of the ribosome to the translocon, the SRP and its receptor disassociate from the complex, allowing resumption of translation (113). This process also seals the pore on the cytosolic side of the ER membrane, releasing BiP from the lumenal side.

## PROCESSING IN THE ENDOPLASMIC RETICULUM

### Signal Peptidase

During translocation, the signal peptide is cleaved from the propeptide by signal peptidase, an integral membrane protein complex on the lumenal surface of the ER. Signal peptidase has been purified from the dog pancreas as a complex of 5 polypeptides with molecular weights of 12, 18, 21, 22/23, and 25 kDa (114,115). The enzyme in the hen oviduct has only 2 subunits of 19 and 22/23kDa (116). The canine and hen 22/23-kDa proteins are glycosylated, and their amino acid sequences are 90% identical (117,118). cDNA encoding the canine 18 (119) and 21 kDa (120) are homologous to 2 yeast SEC11 proteins (121) that are components of the yeast signal peptidase, which contains 4 proteins in total (with molecular weights of 13, 18, 21, and 25 kDa). The 21-kDa protein is absolutely required for enzymatic function in yeast (122,123), raising the question of the exact function of the other proteins. As is the case with processing enzymes, there are great similarities between the yeast and mammalian enzymes (124). Although it appears that the structure of eukaryotic signal peptidases are phylogenetically conserved, the *Escherichia coli* signal peptidase consists of only a single subunit of 323 amino acids (125). Nevertheless, there is some sequence homology between bacterial signal peptidases and subunits of the eukaryotic enzyme. Furthermore, the substrate specificity of the eukaryotic and prokaryotic signal peptidases is similar (126). Eukaryotic signal peptidase has a broad pH optimum and requires phosphatidyl choline as a cofactor (127,128).

Determination of the amino acid sequences that define the substrate specificity of signal peptidase has been difficult because of the enormous structural diversity of signal peptides (129,130). However, there is clearly a hierarchy of preferred substrates for amino acids located at the carboxyl terminus of the signal peptide as follows: Ala>Cys>Gly>Ser>>Thr>Pro>Asn>>Val, Ile, Leu, Tyr, His, Arg, Asp (65,131). Mutations of the signal peptide that increase the number of amino acids between the end of the central hydrophobic domain and the site of cleavage and mutations in the positively charged amino-terminal domain inhibit the cleavage reaction (132–134).

### Disulfide Bond Formation

After peptide prohormones are translocated into the ER lumen, they can undergo intermolecular or intramolecular disulfide bond formation (e.g., proinsulin) (Fig. 2-4). In the case of proinsulin, disulfide bonds are formed before cleavage of proinsulin into its component A and B fragments by removal of the C peptide. Thus, the disulfide bonds that are intramolecular on the prohormone are subsequently converted to intermolecular linkages that cannot be recreated easily after they are reduced. Although spontaneous formation of disulfide bonds of peptides such as somatostatin can occur *in vitro* over a few hours under optimal conditions, *in vivo,* the process occurs either cotranslationally or within seconds after translocation (135). The rapidity of this process suggests that it is catalyzed by an enzyme, the prime candidate being protein disulfide isomerase (PDI) (136–138). In solution, PDI exists as a homodimer ($2 \times 57$ kDa) with a highly acidic isoelectric point (pI) (139). PDI has a broad substrate specificity encompassing relatively small proteins such as insulin, as well as large multidomain proteins such as immunoglobulins (137). PDI also forms the β subunit of a tetrameric enzyme ($\alpha_2\beta_2$) denoted as prolyl-4-hydroxylase, which is responsible for hydroxylation of proline in the formation of procollagen (140). In tissues requiring both PDI and prolyl-4-hydroxylase activities, it appears that the β subunits of prolyl-4-hydroxylase are synthesized in large excess with a fraction being recruited into the prolyl-4-hydroxylase tetramers and the remainder as functional PDI homodimers (140).

Previously, it was thought that glutathione provided the oxidizing equivalents for PDI (141,142). More recently, investigators have identified an ER membrane protein

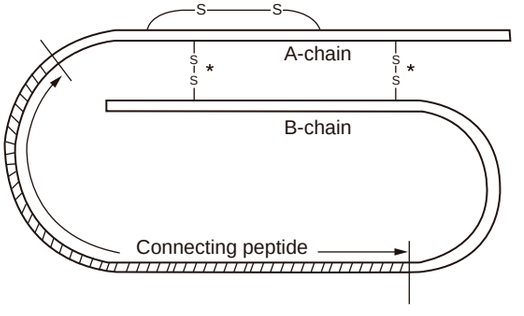



**FIG. 2-4.** Structure of proinsulin. A and B peptides are linked by intramolecular disulfide bonds *(asterisk),* but after the connecting peptide is removed by endoproteolytic cleavage, disulfide bonds are intermolecular. On exocytosis, insulin (A and B peptides) is coreleased with the C peptide.

(Ero 1p) in yeast (143,144) that serves this function. Indeed, Ero 1p directly oxidizes PDI through disulfide exchange (141,145). The reoxidation of Ero 1 involves flavin adenine dinucleotide (FAD) (145).

### Asparagine-Linked N-glycosylation

There are few examples of gastrointestinal peptides with N-linked glycosylation. The primary amino acid sequence of -Asn-X-Thr/Ser, where *X* can be any amino acid except proline, is obligatory for N-glycosylation of asparagine (146). The anterior pituitary glycoprotein hormone family (thyroid-stimulating hormone, follicle-stimulating hormone, and leuteinizing hormone) is the best example of glycosylated hormones. These are dimeric proteins with a common β subunit and different but homologous β subunits that confer specific biological activities. The glycosylation of both subunits is important for their correct assembly into dimers (147). Proopiomelanocortin (POMC) (148) and proenkephalin A (149) are also glycosylated, although the functional significance of this modification is unknown in these peptides. Secretogranin I (also known as chromogranin B) has a single glycosylation site, but it is uncertain whether it is glycosylated *in vivo* (150).

### Protein N-myristoylation

Protein N-myristoylation refers to the cotranslational linkage of myristic acid (C14:0) to the amino-terminal glycine of proteins; protein N-myristoylation is reviewed elsewhere (151). There are no known examples of myristoylated prohormones; however, this modification may play a role in the regulation of a variety of cellular events including posttranslational processing. Examples of N-myristoylated proteins include GTP-binding proteins and the catalytic subunit of cyclic 3′,5′-adenosine monophosphate-dependent protein kinase A.

### Protein Folding

Polypeptides must be folded into a conformation that is compatible with exit from the ER (109,152). Misfolded proteins are tightly but noncovalently bound to a heavy chain binding protein or BiP and retained in the ER (153) until folding is complete and the polypeptide is released on hydrolysis of ATP (154,155). BiP, a member of the heat shock family of proteins (HSP70), binds newly translated and translocated aliphatic single polypeptides and prevents them from folding prematurely (156,157). It is currently unknown whether BiP or other folding proteins (158) are involved in the posttranslational processing of mammalian gastrointestinal prohormones, although BiP is clearly important in the translocation and folding of the yeast prohormone, pro-α factor in the ER (159). Another important folding chaperone is calnexin, but it interacts only with N-glycosylated proteins (158). Because few prohormones are glycosylated, it is hypothesized that this pathway is not involved in prohormone processing. An important factor that should not be forgotten is the role that disulfide bond formation plays in maintaining the folded nature of many polypeptides such as proinsulin.

## TRANSPORT FROM THE ENDOPLASMIC RETICULUM AND THROUGH THE GOLGI

The mechanisms responsible for protein sorting beyond the ER have been the subject of much investigation. Unlike the well-defined sorting of prohormones to the ER lumen through a signal peptide, there is no single unifying mechanism of prohormone transport from the ER and through the Golgi. Two types of sorting mechanisms have been hypothesized. The first is that prohormones are transported in the nonspecific "bulk flow" of contents from the ER to the Golgi in transport vesicles. An alternative hypothesis is that there is some signal contained in the prohormone structure that specifically directs their sorting through the intracellular compartments. This latter hypothesis is the case for resident soluble ER proteins such as BiP and PDI. Investigators noted in the structures of BiP and PDI a carboxyl-terminal consensus sequence KDEL (LysAspGluLeu) (160). Truncated forms of BiP lacking the KDEL sequence are not retained in the ER, but rather are secreted constitutively. In analogous fashion, prohormones destined for secretion but tagged with KDEL are retained in the ER in an unprocessed form (161). The homologous tetrapeptides DKEL, RDEL, and KNEL are all capable of directing ER retention in mammalian cells, whereas the HDEL sequence is used primarily in yeast (162,163). Although the KDEL-tagged proteins could be retained by a KDEL receptor in the ER membrane, it appears that these proteins initially exit the ER and are then recaptured in a salvage compartment at or near the cis-Golgi and returned to the ER (164). A mutant strain of yeast ERD2 (for *E*R *r*etention *d*efective) has been shown to have a defect in the KDEL/HDEL receptor (165). The structure of the *ERD2* gene was then used to aid in the search for a mammalian homologue (166). This powerful technique of identifying genes of fundamental importance to the sorting of proteins in yeast and then using the yeast model to identify a mammalian homologue has been a fruitful approach in the study of peptide hormone processing. The *ERD2* gene encodes a protein of 26 kDa that contains 7 membrane-spanning regions and is highly homologous to a putative human *ERD2*-like gene (167). The mammalian KDEL receptor cycles from the ER to the Golgi and back to the ER, thus retaining lumenal ER proteins within that compartment (168).

Prohormones proceed from the ER to the Golgi stack where they undergo further posttranslation modification. Prohormone movement through the Golgi stack is by bulk flow (169,170) rather than a process mediated by a sorting signal. Bulk transport of soluble ER proteins to the Golgi and through the various Golgi compartments (cis, stack, and

trans-Golgi network [TGN]) was once thought to occur through transport vesicles (171). In this model, the Golgi was a series of stable, disconnected stacks through which proteins were progressively sorted, modified, or "distilled" toward their final destination. Although this model was attractively simple, it now appears that the Golgi is a much more fluid organelle (171–174).

Although the transport vesicle model was indeed attractive, the described vesicles were too small (70 nm in diameter) to transport many secreted proteins. An alternative model is that newly synthesized proteins move from the ER to the cis-Golgi cisternae located near the ER. This newly formed cisternae progresses through the Golgi stack from the ER to the trans-Golgi (175). The transport vesicles in this case merely shuttle enzymes that characterize the various layers of the Golgi back through the cisternae (176–179). Thus, prohormones are transported from the ER to the cis-Golgi and are not transported out of this compartment, but rather are carried forward to the trans side as newer enzymes and proteins are added to the cis-Golgi.

The nature of the ER to Golgi transport has been studied extensively (173). The ER membrane has a fixed number of exit sites from which proteins leave the lumen (180). The ER membranes cause buds that eventually become coated on their outer cytoplasmic surface with dispersed cytoplasmic proteins (coatamers, coat promoter, or COPs) (181–184). The budding ER vesicle is coated with COPII and traps prohormones together with other ER proteins (185–187). The COPII-coated vesicles then uncoat and fuse into a larger vesicular tubular complex (VTC) (185). It appears that the VTC is not continuous with the ER membrane. Eventually, the VTC combines with COPI (188). After fusing with the Golgi membrane (in a GTP-dependent manner) (189–191), the COPI-associated VTCs return ER proteins for recycling. Thus, nonhydrolyzable analogues of GTP such as GTP-γ-s interfere with fusion and block transport through the stack (192). VTCs move from the ER to the Golgi along a microtubular network that is powered by the dynein–dynactin motor (190,193). The fungal metabolite Brefeldin A, which is known to block protein transport through the Golgi stack, blocks the binding of the coatamer complex to budding Golgi membranes (194–197). The finding that Brefeldin A interferes with the posttranslational processing of progastrin suggests that this pathway is involved in the sorting of prohormones, as well as other soluble secretory proteins (198,199).

Proteins secreted via either the constitutive or regulated secretory pathways share a common trail from the ER through the Golgi stack, but they diverge in the TGN where proteins are sorted according to their final destination (169,200–202). The sorting signal for enzymes destined for lysosomes involves a glycosylation reaction that occurs in the Golgi stack to attach mannose-6-phosphate residue proteins. A receptor protein in the TGN specifically binds mannose-6-phosphate–modified proteins (203,204) and directs their sorting to lysosomes. To date, searches for a common sequence (KDEL-like) or posttranslational modifications (mannose-6-phosphate–like) in the structure of prohormones that might direct sorting to secretory vesicles in the TGN have not been successful.

Although investigators have long sought to elucidate "the" Golgi sorting signal in neuroendocrine cells, none has been entirely successful. Indeed, it appears that three different mechanisms may be responsible for prohormone sorting to secretory vesicles. These include sorting signal motifs, aggregation, and membrane or lipid raft binding (205). Initially, the search for a sorting signal was pursued vigorously. In a fashion akin to the signal peptide ("pre" region of preprohormones), investigators sought sequences in the "pro" region of prohormones that, although lacking homology in their primary amino acid sequence, still contain sufficient structural information to direct sorting in the TGN. An α-helical motif with three leucine residues occupying one side of the helix was proposed as such a signal, but this hypothesis was not proved (206). In other studies, a chimeric protein containing the "prepro" region of somatostatin at the amino terminus and a constitutively secreted protein such as γ-globulin at the carboxyl terminus were sorted and processed in the secretory pathway (207). Studies with POMC and somatostatin precursors containing deletions in the "pro" region indicate the presence of sorting information at these sites, as well as in other portions of the peptide (208,209). In contrast, deletion of the "pro" sequence from trypsinogen and renin did not disrupt the routing of these proteins into the secretory pathway (210,211). A study expressed neuropeptide Y (NPY) fragments tagged with green fluorescent protein (GFP). GFP, a jellyfish protein not normally secreted, was correctly sorted, stored, and released from neuroendocrine cells when fused to half of the prepro-NPY sequence or only the signal sequence alone of pre-NPY (212). Thus, it appears that some prohormones are likely sorted by a specific signal found in their "pro" regions, but this does not appear to be a universal finding.

A second sorting hypothesis is selective aggregation (213–215) of prohormones into acidic clathrin-coated secretory vesicles in the presence of high concentrations of divalent cations such as $Zn^{2+}$ or $Ca^{2+}$. Support for this hypothesis comes from observations that specific mutations in the structure of proinsulin that result in inhibition of hexamer formation with zinc also impede processing (48,216). Furthermore, *in vitro* studies have demonstrated that intravesicular conditions (pH 5.2 and 10 mM $Ca^{2+}$) can result in selective precipitation of peptides that exit the cell through the regulated pathology of secretion. This applies to secreted proteins such as secretogranin II but not proteins that are constitutively secreted such as immunoglobulins (217). A heterodimeric protein in adrenal chromaffin granules, termed *glycoprotein III,* can selectively aggregate with two prohormone-processing enzymes carboxypeptidase E (CPE) and a dibasic endoprotease (218). Evidence contradictory to the selective aggregation hypothesis can be found in studies with guinea pig proinsulin, which does not form hexamers with zinc and yet is sorted and processed with high efficiency (219). In addition, in the marine mollusk Aplysia, the egg-laying hormone precursor is processed into two distinct mature hormone products that are sorted into different