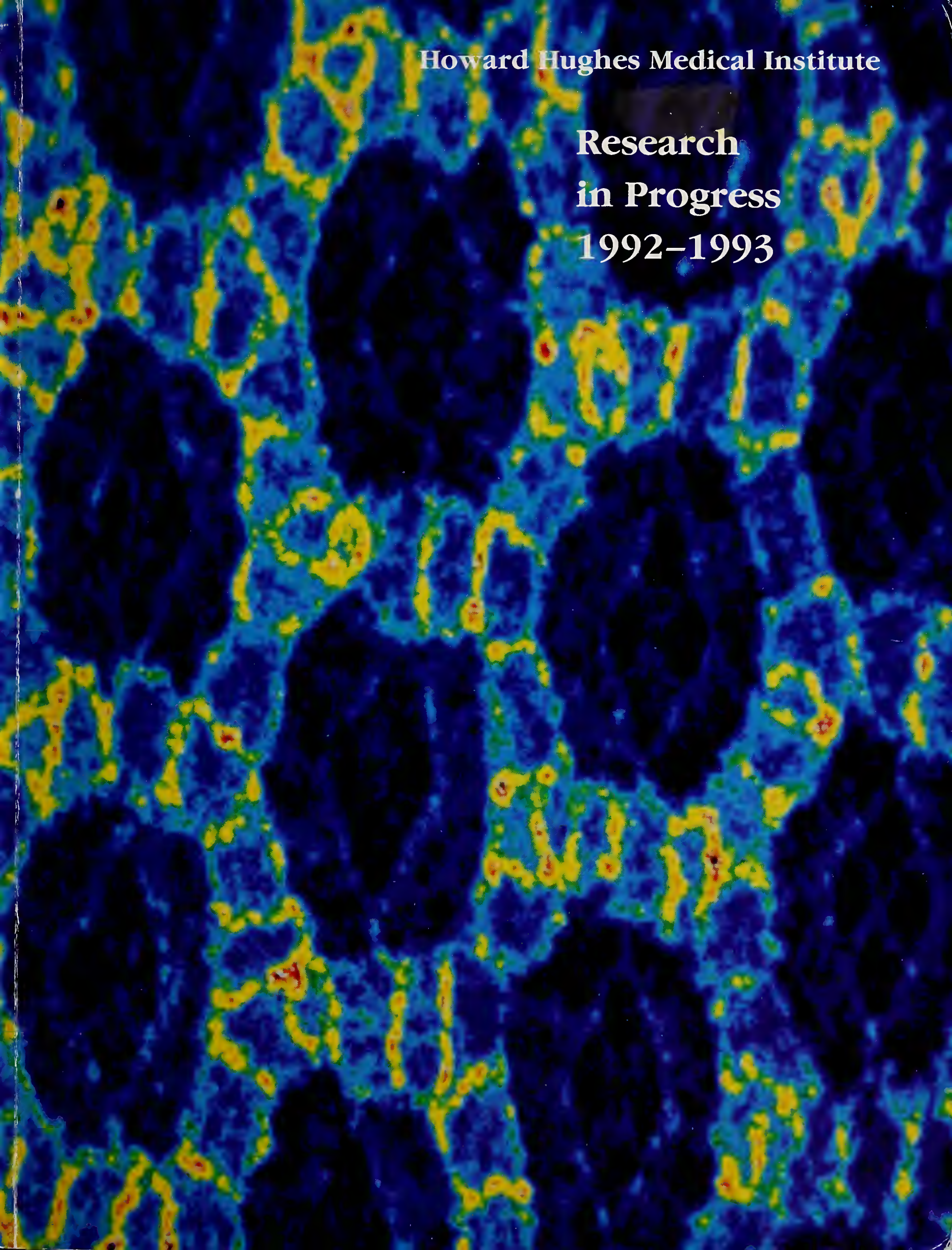
Howard Hughes Medical Institute

Research in Progress 1992–1993

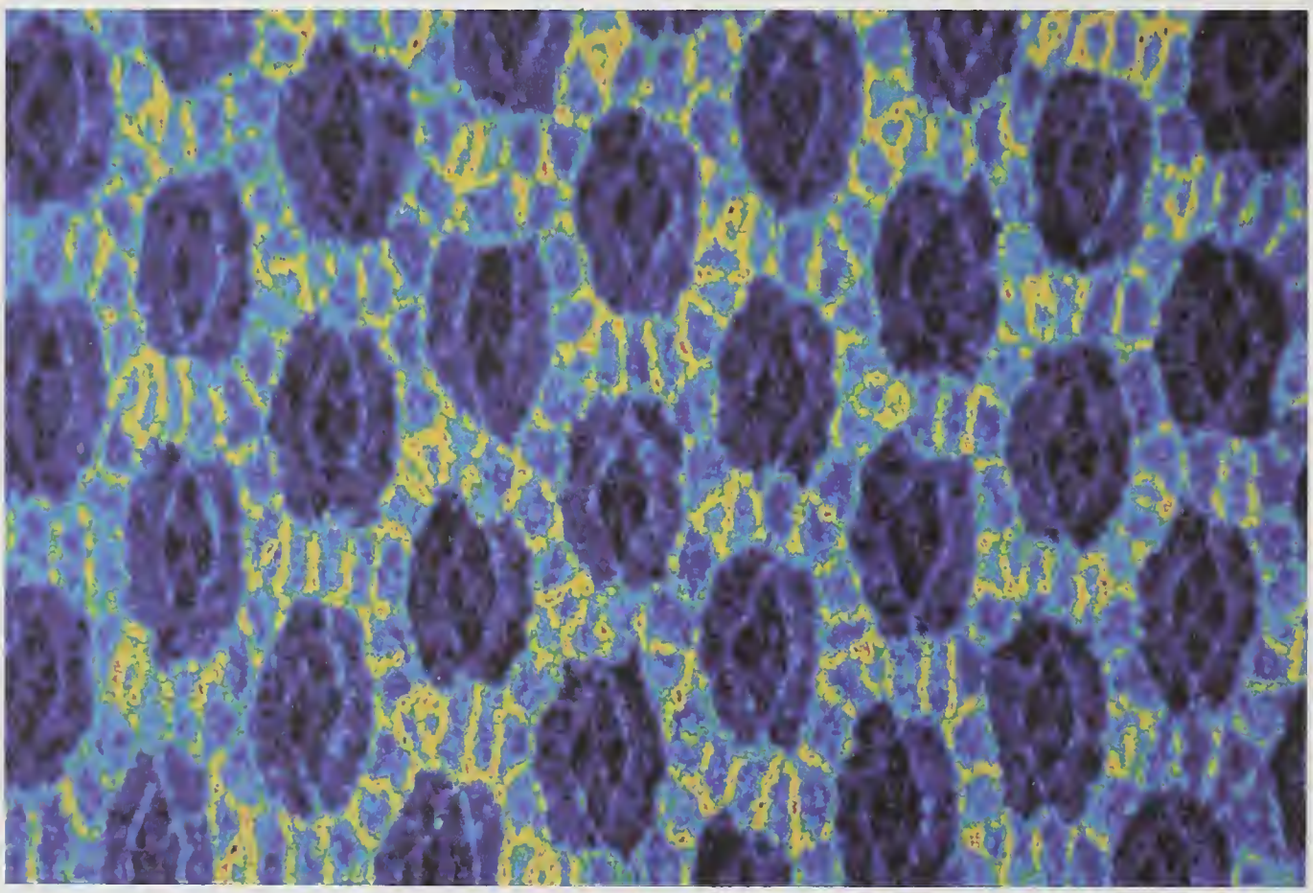


Research in Progress
in the Howard Hughes Medical Institute, 1992–1993

Founded in 1953 and incorporated in Delaware, the Howard Hughes Medical Institute is a nonprofit scientific and philanthropic organization. It is dedicated to the promotion of human knowledge within the basic sciences, principally medical research and medical education, and the use of such knowledge for the benefit of humanity. The Institute is directly engaged in medical research and is qualified as a medical research organization under the U.S. tax code. It is an equal opportunity employer.

Howard Hughes Medical Institute
6701 Rockledge Drive
Bethesda, Maryland 20817
(301) 571-0200

After February 8, 1993:
Howard Hughes Medical Institute
4000 Jones Bridge Road
Chevy Chase, Maryland 20815-6789
(301) 215-5000



Developing eye of the common fruit fly Drosophila melanogaster, *as imaged by a computer-based confocal microscope. Red and yellow represent the highest concentrations of Notch, a cell-surface protein that is necessary for the cell-to-cell interactions that decide the developmental fate of each cell within the eye.*

From Fehon, R.G., Johansen, K., Rebay, I., and Artavanis-Tsakonas, S. 1991. J Cell Biol *113:657–669. Copyright permission by the Rockefeller University Press.*

Printed in the United States of America

Howard Hughes Medical Institute

Research in Progress 1992–1993

Contents

Foreword

THE Howard Hughes Medical Institute (HHMI) presents the seventh volume of *Research in Progress*, a series begun in 1986 to describe the research activities of the Institute. Early in 1992 the scientists associated with HHMI at that time were asked to provide statements in nontechnical terms of their past accomplishments, current work, and plans for the future, in order to provide a concise but informative picture of the Institute's research in progress. You will read their personal essays in the pages that follow. We are pleased that this annual publication has come to be a useful source of information for scientists and nonscientists interested in biomedical research, as well as for the members of the HHMI family.

The Institute also publishes as a companion volume the formal *Annual Scientific Report*. This is the official archival record of the research of each HHMI laboratory and includes yearly bibliographies as well as descriptions of other HHMI scientific activities. A general *Annual Report of the Howard Hughes Medical Institute* describes the various programs for a lay audience and gives a summary of the Institute's financial data. With the recent publication of *From Egg to Adult*, which describes the exciting process of biological development, the Institute has expanded its series of reports on subjects of current scientific interest written for a general audience. This booklet joins the widely acclaimed *Finding the Critical Shapes* and *Blazing a Genetic Trail* that have been in great demand by a broad spectrum of readers, especially teachers who have used these reports in classrooms across this country and abroad.

In recognition of the fact that the boundaries of biomedical science are not constrained by national borders, the Institute launched its International Research Scholars Program in the spring of 1991. This is a limited and experimental program with the purpose of providing research support in selected countries for promising scientists whose careers are in a developing phase but who have already made significant contributions to biomedical research. The international program is separate from the Institute's ongoing medical research program, described in the first section of this volume, in which scientists join HHMI as full-time employee-investigators and are supported through the Institute's operation as a Medical Research Organization (MRO). The Scholars do not join the HHMI staff, but instead each receives research support through a grant.

For the initial awards in this program, selected scientists in the countries that are our immediate neighbors, Canada and Mexico, were invited to compete. Twenty-four outstanding investigators received five-year awards (14 from Canada and 10 from Mexico). You will read their descriptions of the research in their laboratories in the second section of this volume. For awards in 1993, the Institute has turned to scientists in the United Kingdom, Australia, and New Zealand. Those selected for grant support will be announced in late 1992, and their work will be included in the next volume of *Research in Progress*.

The Institute has a large grants program that complements its MRO activities. In addition to the International Research Scholars initiative, the grants program supports education in the biomedical and related sciences at the precollege, college undergraduate, graduate, and postdoctoral levels. New in 1992 were grants awarded to certain science museums, science and technology centers, children's museums, and natural history museums to assist in education and outreach programs for elementary school children and their families. The Institute's annual publication *Grants for Science Education* details these initiatives.

A very important event for HHMI occurred in the spring of 1991, when construction began on the buildings that will become our new headquarters and conference center complex in Chevy Chase, Maryland. At this writing, the construction is entering its final phases, and the attractiveness of the buildings of this large facility and how they fit into the gentle hills and valleys of the lovely site are exceeding our original expectations. We expect to occupy the buildings in early 1993. The facility provides the administrative focus for the Institute's 222 investigators (as of July 1, 1992) whose laboratories are located at 53 institutions across this country.

We invite you to share the excitement of the research of the HHMI investigators whose work forms the core of the Institute's activities and of the studies of our International

Research Scholars that have further enriched our scientific program. We are especially pleased that the Institute's research on problems of medical concern to developing countries is extended through the basic studies of a number of these Scholars. The advances toward understanding biological processes and disease mechanisms that are described herein move us further toward our goal of the betterment of the human condition.

Purnell W. Choppin, M.D.
President

Trustees, Officers, and Principal Staff Members

Trustees

Alexander G. Bearn, M.D.
Adjunct Professor
The Rockefeller University
Professor Emeritus of Medicine
Cornell University Medical College
Former Senior Vice President
Merck Sharp & Dohme, International

Helen K. Copley
Chairman of the Corporation and
Chief Executive Officer
The Copley Press, Inc.

Frank William Gay
Former President and Chief
Executive Officer
SUMMA Corporation

James H. Gilliam, Jr., Esq.
Executive Vice President
Beneficial Corporation

Hanna H. Gray, Ph.D.
President
The University of Chicago

William R. Lummis, Esq.
Chairman of the Board of Directors
SUMMA Corporation

Irving S. Shapiro, Esq., *Chairman*
Of Counsel
Skadden, Arps, Slate, Meagher & Flom
Former Chairman and Chief
Executive Officer
E.I. du Pont de Nemours and Company

George W. Thorn, M.D.,
Chairman Emeritus
Professor Emeritus
Harvard Medical School

James D. Wolfensohn
President
James D. Wolfensohn Incorporated

Officers

Purnell W. Choppin, M.D.
President

W. Maxwell Cowan, M.D., Ph.D.
Vice President and Chief Scientific Officer

Graham O. Harrison
Vice President and Chief Investment Officer

Joseph G. Perpich, M.D., J.D.
Vice President for Grants and
Special Programs

José E. Trías, Esq.
Vice President and General Counsel

Robert C. White
Vice President and Chief Financial Officer

Principal Staff Members

Craig A. Alexander, Esq.
Associate General Counsel

Stephen A. Barkanic
Grants Program Officer

W. Emmett Barkley, Ph.D.
Director of Laboratory Safety

Lillian H. Blucher
Managing Director—Investments

Winfred J. Clingenpeel
Director of Purchasing

David Davis-Van Atta
Grants Program Officer

Barbara Filner, Ph.D.
Grants Program Officer

Patricia S. Gage, Esq.
Associate General Counsel

James R. Gavin III, M.D., Ph.D.
Senior Scientific Officer

Donald H. Harter, M.D.
Senior Scientific Officer and
Director, HHMI-NIH Research
Scholars Program

Patricia J. Hoben, Ph.D.
Grants Program Officer
(*through September 4, 1992*)

John A. Jones
Director of Computer Services

David W. Kingsbury, M.D.
Senior Scientific Officer

Laura A. Kumin, Esq.
Associate General Counsel

Joan S. Leonard, Esq.
Associate General Counsel

Robert H. McGhee
Director of Research Facilities Planning

Alan E. Mowbray
Director of Management Services

Robert C. Mullins
Director of Internal Audit

Edward J. Palmerino
Assistant Controller

Robert A. Potter
Director of Communications

Donald C. Powell
Director of Human Resources

Ellen B. Safir
Managing Director—Investments

Mark W. Smith
Controller

Claire H. Winestock, Ph.D.
Senior Scientific Officer

Medical Advisory Board

Michael S. Brown, M.D.
Paul J. Thomas Professor of Medicine and Genetics
Director, Center for Genetic Disease
University of Texas Southwestern Medical Center at Dallas

William A. Catterall, Ph.D.
Professor and Chair
Department of Pharmacology
University of Washington School of Medicine

John E. Dowling, Ph.D.
(through December 31, 1992)
Maria Moors Cabot Professor of Natural Science
Department of Cellular and Developmental Biology
Harvard University

Thomas J. Kelly, Jr., M.D., Ph.D.
(effective January 1, 1993)
Chairman
Department of Molecular Biology and Genetics
The Johns Hopkins University School of Medicine

Hugh O. McDevitt, M.D.
(through December 31, 1992)
Burt and Marion Avery Professor of Immunology
Professor of Microbiology, Immunology, and Medicine
Stanford University School of Medicine

Dinshaw Patel, Ph.D.
(effective January 1, 1993)
Abby Rockefeller Mauzé Chair in Experimental Therapeutics
Member, Cellular Biochemistry and Biophysics Program
Memorial Sloan-Kettering Cancer Center

William E. Paul, M.D.
Chief, Laboratory of Immunology
National Institute of Allergy and Infectious Diseases
National Institutes of Health

Frederic M. Richards, Ph.D.
(through December 31, 1992)
Sterling Professor Emeritus of Molecular Biophysics and Biochemistry
Yale University

Janet D. Rowley, M.D.
Blum-Riese Distinguished Service Professor
Departments of Medicine and Molecular Genetics and Cell Biology
The University of Chicago

David D. Sabatini, M.D., Ph.D.
Frederick L. Ehrman Professor and Chairman
Department of Cell Biology
New York University School of Medicine

Phillip A. Sharp, Ph.D.
John D. MacArthur Professor of Biology and Head, Department of Biology
Massachusetts Institute of Technology

Melvin I. Simon, Ph.D.
(effective January 1, 1993)
Biaggini Professor of Biology
California Institute of Technology

William S. Sly, M.D.
(through December 31, 1992)
Alice A. Doisy Professor and Chairman
Edward A. Doisy Department of Biochemistry and Molecular Biology
St. Louis University School of Medicine

Lloyd H. Smith, Jr., M.D., *Chairman*
Professor of Medicine and Associate Dean
University of California, San Francisco, School of Medicine

Jonathan W. Uhr, M.D.
(effective January 1, 1993)
Professor and Chairman
Department of Microbiology
University of Texas Southwestern Medical Center at Dallas

Scientific Review Board

Cell Biology and Regulation

Alfred G. Gilman, M.D., Ph.D.
Professor and Chairman
Department of Pharmacology
University of Texas Southwestern Medical Center at Dallas

Tony Hunter, Ph.D.
Professor of Molecular Biology and Virology
The Salk Institute

Stuart Kornfeld, M.D.
Professor of Medicine
Washington University School of Medicine

David J. L. Luck, M.D., Ph.D.
Professor of Cell Biology
The Rockefeller University

Thomas D. Pollard, M.D.
Professor and Director
Department of Cell Biology and Anatomy
The Johns Hopkins University School of Medicine

Lucille Shapiro, Ph.D.
Joseph D. Grant Professor in the School of Medicine and Chairman, Department of Developmental Biology
Stanford University School of Medicine

Genetics

David Botstein, Ph.D.
Professor and Chairman
Department of Genetics
Stanford University School of Medicine

Ira Herskowitz, Ph.D.
Professor and Chairman
Department of Biochemistry and Biophysics
Head, Division of Genetics
University of California, San Francisco, School of Medicine

Anthony P. Mahowald, Ph.D.
Louis Block Professor and Chairman
Department of Molecular Genetics and Cell Biology
The University of Chicago

Thomas P. Maniatis, Ph.D.
Professor of Biochemistry and Molecular Biology
Harvard University

Malcolm A. Martin, M.D.
Chief, Laboratory of Molecular Microbiology
National Institute of Allergy and Infectious Diseases
National Institutes of Health

Carolyn W. Slayman, Ph.D.
Professor and Chairman
Department of Genetics
Professor, Department of Cellular and Molecular Physiology
Yale University School of Medicine

Harold E. Varmus, M.D.
American Cancer Society Professor of Molecular Virology
Department of Microbiology and Immunology
University of California, San Francisco, School of Medicine

Immunology

Joseph M. Davie, M.D.
(through June 30, 1992)
President
G.D. Searle and Company

Matthew D. Scharff, M.D.
Professor of Cell Biology
Albert Einstein College of Medicine

Ursula Storb, M.D.
(effective January 1, 1993)
Professor
Department of Molecular Genetics and Cell Biology
The University of Chicago

Emil R. Unanue, M.D.
Professor and Chairman
Department of Pathology
Washington University School of Medicine

Ellen S. Vitetta, Ph.D.
Professor of Microbiology and Director,
Cancer Immunobiology Center
University of Texas Southwestern Medical
Center at Dallas

Neuroscience

Floyd E. Bloom, M.D.
Chairman, Department of
Neuropharmacology
Member, The Scripps Research Institute

Arthur M. Brown, M.D., Ph.D.
Professor and Chairman
Department of Molecular Physiology
and Biophysics
Baylor College of Medicine

Zach W. Hall, Ph.D.
Professor and Chairman
Department of Physiology
University of California, San Francisco,
School of Medicine

Stephen F. Heinemann, Ph.D.
(effective January 1, 1993)
Faculty Chairman
Director of Molecular Neurobiology
Laboratory
The Salk Institute

Carla J. Schatz, Ph.D.
(effective July 1, 1993)
Professor of Neurobiology
Department of Molecular and Cell Biology
University of California, Berkeley

Melvin I. Simon, Ph.D.
(through December 31, 1992)
Biaggini Professor of Biology
California Institute of Technology

Structural Biology

David R. Davies, Ph.D.
Chief, Section on Molecular Structure
Laboratory of Molecular Biology
National Institute of Diabetes and Digestive
and Kidney Diseases
National Institutes of Health

Michael N. G. James, D.Phil.
Professor of Biochemistry
University of Alberta

Dinshaw Patel, Ph.D.
(through December 31, 1992)
Abby Rockefeller Mauzé Chair in
Experimental Therapeutics
Member, Cellular Biochemistry
and Biophysics Program
Memorial Sloan-Kettering Cancer Center

Michael G. Rossmann, Ph.D.
Hanley Professor of Biological Sciences
Purdue University

Peter E. Wright, Ph.D.
(effective January 1, 1993)
Member and Chairman
Department of Molecular Biology
The Scripps Research Institute

Locations of Howard Hughes Medical Institute Laboratories

Alabama	University of Alabama at Birmingham and associated hospitals
California	California Institute of Technology and associated hospitals, Pasadena The Salk Institute for Biological Studies, San Diego Stanford University and the Stanford University Hospital, Palo Alto University of California, Berkeley, and associated hospitals University of California, Los Angeles, and associated hospitals University of California, San Diego, and the UCSD Medical Center University of California, San Francisco, and associated hospitals University of Southern California, Los Angeles, and associated hospitals
Colorado	National Jewish Center for Immunology and Respiratory Medicine, Denver University of Colorado at Boulder and the University's Health Sciences Center University of Colorado Health Sciences Center, Denver, and associated hospitals
Connecticut	Yale University and associated hospitals, New Haven
Georgia	Emory University School of Medicine, Atlanta, and associated hospitals
Illinois	Northwestern University and associated hospitals, Evanston The University of Chicago and The University of Chicago Hospitals
Indiana	Indiana University, Bloomington, and associated hospitals Indiana University School of Medicine, Indianapolis, and associated hospitals
Iowa	The University of Iowa and associated hospitals, Iowa City
Maryland	The Carnegie Institution of Washington, Baltimore, and The Johns Hopkins Hospital The Johns Hopkins University and Hospital, Baltimore
Massachusetts	Brandeis University, Waltham, and associated hospitals Brigham and Women's Hospital, Boston The Children's Hospital, Boston Harvard College, Arts and Sciences, Cambridge Harvard Medical School, Boston Massachusetts General Hospital, Boston Massachusetts Institute of Technology and associated hospitals, Cambridge Tufts University School of Medicine and associated hospitals, Boston University of Massachusetts, Worcester, and associated hospitals
Michigan	University of Michigan and associated hospitals, Ann Arbor
Missouri	Washington University and associated hospitals, St. Louis
New Jersey	Princeton University and associated medical centers, Princeton
New York	Albert Einstein College of Medicine of Yeshiva University, Bronx, and associated hospitals Cold Spring Harbor Laboratory and associated hospitals, Cold Spring Harbor Columbia University and associated hospitals, New York City Cornell University Medical College, New York City Memorial Sloan-Kettering Cancer Center, New York City New York University (Medical Center and Washington Square) and associated hospitals, New York City The Rockefeller University and Rockefeller University Hospital, New York City State University of New York at Stony Brook and University Hospital at Stony Brook
North Carolina	Duke University, including Duke University Medical Center, Durham

Oklahoma	Oklahoma Medical Research Foundation and associated hospitals, Oklahoma City
Oregon	University of Oregon and associated hospitals, Eugene
Pennsylvania	University of Pennsylvania and associated hospitals, Philadelphia
Tennessee	St. Jude Children's Research Hospital, Memphis Vanderbilt University, including Vanderbilt University Hospital, Nashville
Texas	Baylor College of Medicine and associated hospitals, Houston Rice University and associated hospitals, Houston University of Texas Southwestern Medical Center at Dallas and associated hospitals
Utah	University of Utah, including University of Utah Medical Center, Salt Lake City
Washington	Fred Hutchinson Cancer Research Center, Seattle University of Washington and associated hospitals, Seattle
Wisconsin	University of Wisconsin–Madison, and associated hospitals

Other Institute Facilities

Maryland	HHMI-NIH Cloister at the Mary Woodard Lasker Center for Science and Education on the NIH campus, Bethesda
New York	Synchrotron Beam Lines, Brookhaven National Laboratory (under construction)

Introduction

THE annual publication of *Research in Progress* has become one of the highlights of the Howard Hughes Medical Institute's research activities. By providing a synoptic overview of the research being conducted by the Institute's investigators, in essentially nontechnical language, it has come to be greatly appreciated not only by those interested in the Institute's activities as a Medical Research Organization, but by many others who view it as an accessible introduction to current biomedical research.

The present volume in this series follows fairly closely the pattern that has evolved over the past three or four years: it provides a snapshot of the status of the Institute's research in early 1992. As in the previous volume, this issue includes a series of introductory essays that are intended to assist the reader who has had little or no background in biology or medicine. These essays are not intended to be exhaustive nor simply to replicate material readily available in most standard biology or medical texts; rather they are intended to serve as an expanded glossary of terms, defining in a general way many of the terms used in the individual reports. To emphasize this feature of the essays, these technical terms are shown in bold type. The essays are based on an initial set of drafts prepared for each of the five programmatic areas in which the Institute conducts research: cell biology and regulation; genetics; immunology; neuroscience; and structural biology. The original drafts were kindly prepared by Drs. Richard O. Hynes, Philip Leder, Charles A. Janeway, Eric R. Kandel, and Stephen C. Harrison, respectively. We are grateful to them and to a number of their colleagues within the Institute who commented on the essays or provided additional information and, in some cases, clarified ambiguities or corrected errors in the original text. As in last year's volume the text of the essays has been supplemented by a number of illustrations, derived for the most part from general texts or scientific papers that are identified in the accompanying legends. We hope that these diagrams and photographs will make the essays even more useful.

For readers who wish to know about specific research being conducted by HHMI investigators, we have provided once again a detailed index. The index lists almost every topic or term, from *abd-A* (a gene that directs the segmental development of the abdomen—first identified in the fruit fly *Drosophila melanogaster*) to the *zif268* gene and, in between, the less arcane subjects of AIDS, cancer, cystic fibrosis, diabetes, hemophilia, muscular dystrophy, and obesity. Several readers have remarked how useful the addition of an index has been, and we are grateful once more to Diana Witt for preparation of the index.

The greater part of the volume consists of a series of short reports by the investigators associated with the Institute during 1992. This portion of the volume is again larger than its predecessor, reflecting the increased number of investigators whose work is represented. In addition, reports prepared by the Institute's International Research Scholars are included in a section following the investigators' reports. These 24 scientists are the first recipients of research grants awarded under the Institute's new international initiative. The reports submitted by each investigator and International Research Scholar have been collated and edited by Dr. Claire H. Winestock, Senior Scientific Officer; Elizabeth Cowley, Copy Editor; William T. Carrigan, Editor/Writer; Gail Markley, Manager of Publications; and Kimberly A. Cornejo, Permissions Editor. Laura North also assisted with the editing of manuscripts. We are grateful to them and to the many investigators who have provided us with suitable illustrations taken from their ongoing research. Not only do the illustrations help to clarify the written text, but they also greatly enhance the aesthetic appeal of the volume.

Cell Biology and Regulation Program

The Cell Biology and Regulation Program is the oldest of the Institute's research programs. Originally referred to as Metabolic Regulation, which reflected the program's roots in clinical studies of metabolic and endocrine disorders, the title of this program was changed six years ago to reflect more accurately its principal theme—the biology of cells, the factors that regulate their normal growth and distinctive functions, and the ways in which cells interact with each other. In this sense the program has come to occupy a critical position in the Institute's research endeavors, and it lies firmly within the mainstream of contemporary biological research.

The human body contains several trillion (i.e., million million) cells of a thousand or more distinct types. Research in the field of cell biology seeks to understand how these various cells are constructed and organized, how they differ from one another, how they sense and respond to outside influences, how they interact with their neighbors to form more complex tissues and organs, and, in general, how the cells of the body are integrated to produce an appropriately functioning organism. Equally important, research in this area is aimed at understanding how these cellular functions are perturbed by disease. To this extent the problems addressed by investigators in the Cell Biology and Regulation Program inevitably impinge on related work in genetics, development, neuroscience, and immunology. For a Medical Research Organization it is especially gratifying to see how many of the insights gleaned from these studies are already beginning to throw light on such medically important problems as diabetes, heart disease, cancer, muscular dystrophy, cystic fibrosis, and a number of other genetic disorders.

The strikingly rapid advances that have occurred in cell biology in recent years have been due, in large part, to earlier progress in biochemistry and cellular physiology, but especially to the dramatic developments that have occurred in molecular biology since the early 1960s. The techniques developed in these fields have been invaluable to cell biology, which has always been quick to apply different and newly emerging approaches to the solution of the many problems of cell structure and function. To understand the types of research being conducted in contemporary cell biology, it may be helpful to begin with a general account of a "typical" animal cell (Figure 1).

Near the center of each cell is the **nucleus**, which contains the **genes** (Figure 2) that are made of DNA and encode the information necessary to construct an entire organism and maintain its day-to-day activities. The entire complement of genes is the **genome**, which comprises a set of instructions encoded in the sequences of the DNA molecules (as described more fully in the section on genetics). The human genome consists of 46 chromosomes—22 pairs of autosomes and 2 sex chromosomes. The two copies of each autosome are inherited from the mother and the father, respectively; in females there are two X chromosomes (one from each parent), while in males the Y chromosome is always inherited from the father and the X chromosome (Figure 3) is always inherited from the mother. The 46 chromosomes comprise a total of about 3 billion pairs of nucleotides. Estimates vary, but it is thought that there may be as many as 100,000 genes in the human genome. These genes vary in length from around 1,000 to about 2 million nucleotides. Each gene encodes the information for a particular cellular structure or function. This information, which can be likened to a computer language, is first read (**transcribed**) into RNA, and the message contained within the nucleotide sequence of the RNA molecule is then decoded (**translated**) by the machinery of the cell into a different language, or chemical structure. While the RNA transcript mirrors exactly the DNA sequence of the gene, the messenger RNA (mRNA) that is translated into the amino acid sequence of the encoded protein is a highly edited message (Figure 4). Generally several intervening (or noncoding) sequences called **introns** are selectively removed from the transcript. Introns account for a considerable proportion of the DNA in all higher organisms and, together with several other noncoding stretches of DNA (spacer DNA, satellite DNA, and other repetitive DNA sequences), are sometimes referred to as **selfish** or **junk** DNA. In humans it is estimated that more than 90 percent of the DNA in the genome is of this kind. The possible functions (if any) of this noncoding DNA are not known, but the mechanisms whereby the coding sequences or **exons** are spliced out and joined together (sometimes in different order—a process known as alternative splicing) is a subject of considerable interest at present. The chemical language of the cell has 20 different characters or units, known as **amino acids**, which are linked together, again in linear arrays, to make **proteins**.

Whereas DNA is the blueprint directing the cell's development and function, proteins are the molecules from which cells are built and which carry out most cellular functions. Most genes encode proteins, and each cell contains about 10,000 different types of protein. That is, each cell uses only about 10 percent of the total set of genes at any one time. This raises two of the central questions in cell biology today: 1) How are genes turned on and off so that each cell type expresses only its appropriate set of genes and contains only its correct complement of proteins? 2) How are the genes in a given cell regulated, so that the cell can respond appropriately to outside influences by changing either the pattern of

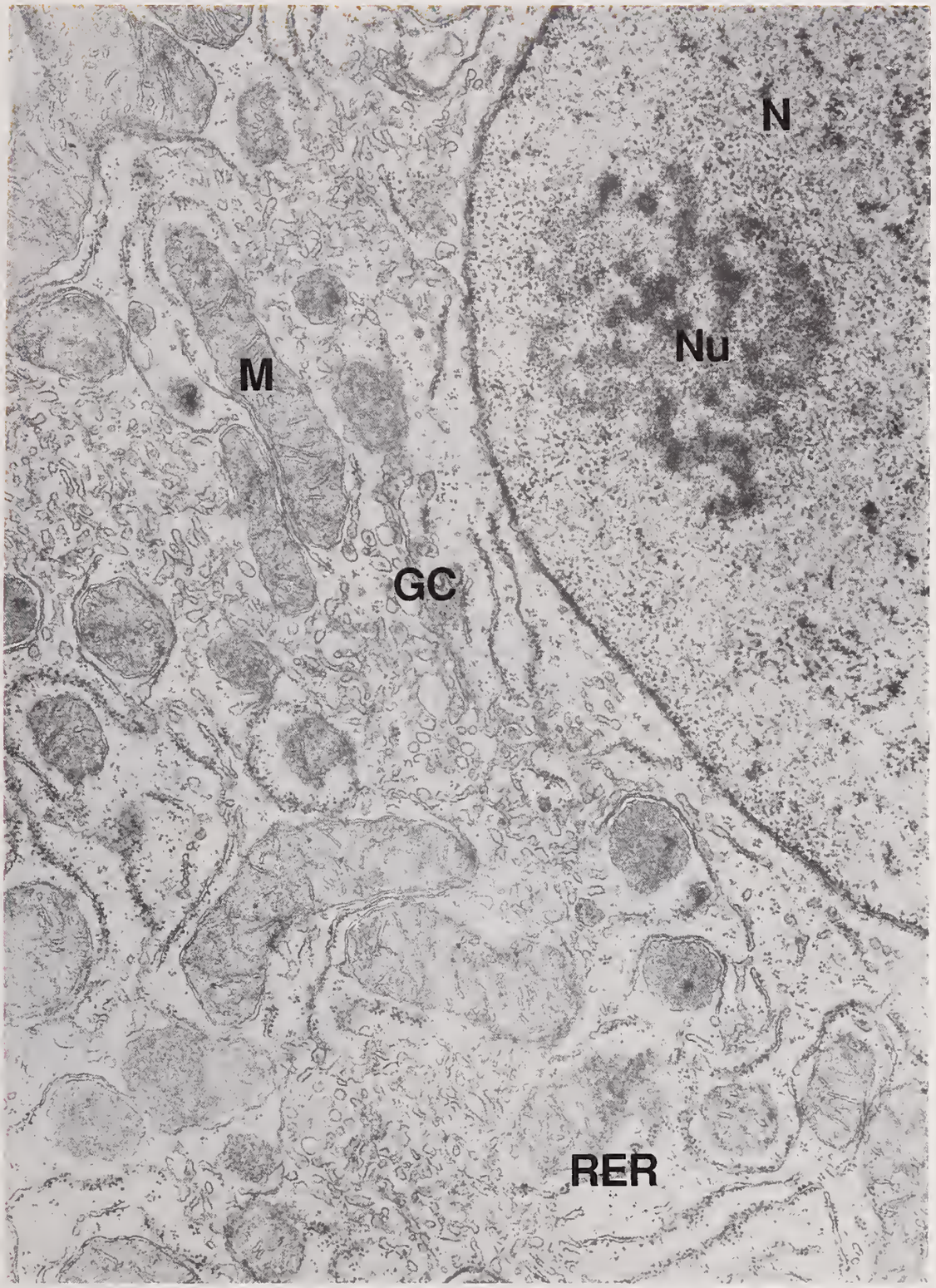


Figure 1. A small section of a typical mammalian cell as seen in the electron microscope. Part of the nucleus (N) with its surrounding membrane and nucleolus (Nu) are shown in the upper right; the cytoplasm that occupies the rest of the electron micrograph contains several different organelles, including several mitochondria (M), parts of the Golgi complex (GC), and the rough endoplasmic reticulum (RER). Magnification approximately 50,000 ×.

Micrograph provided by David D. Sabatini.

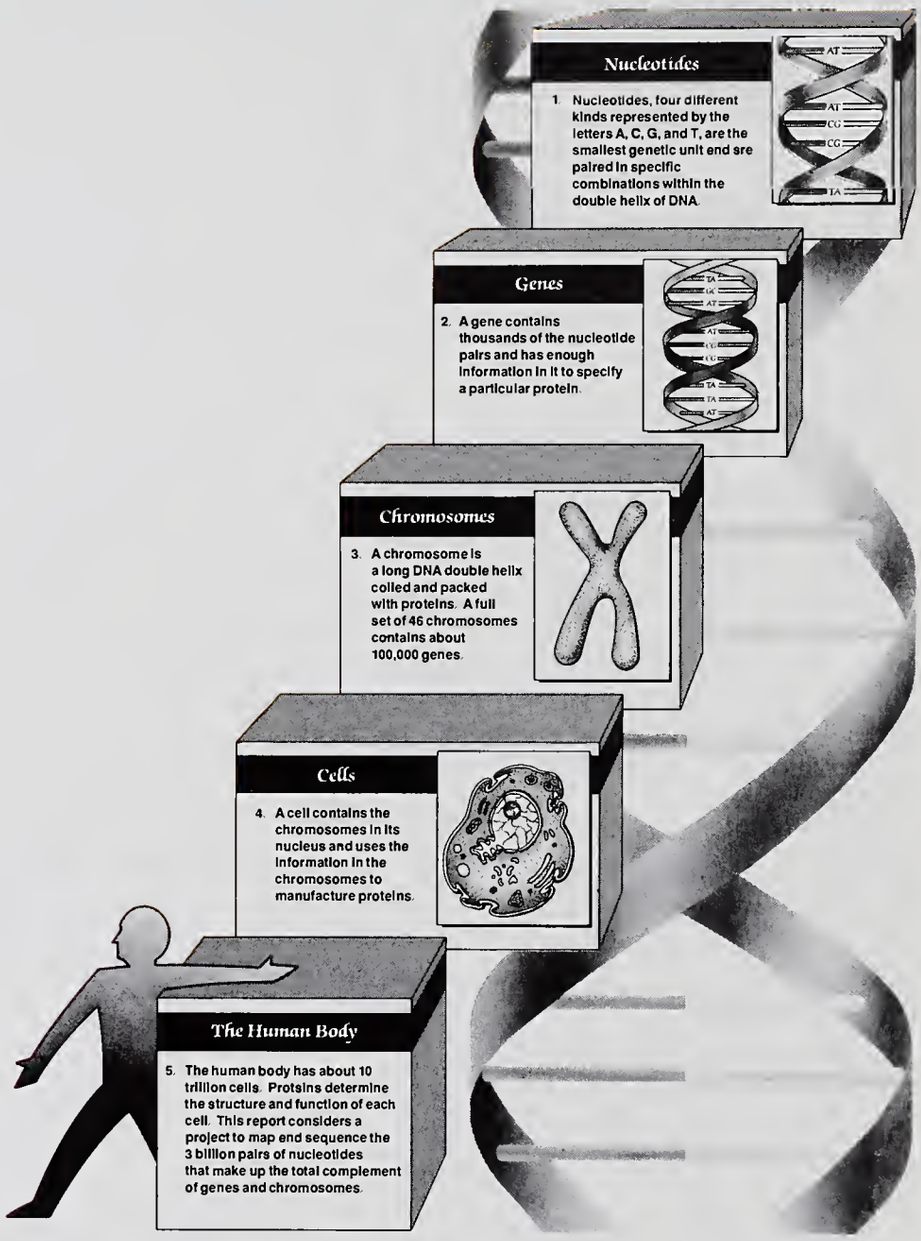


Figure 2. From genes and chromosomes to cells, organs, and entire bodies. Adapted from an illustration by Warren Isensee for The Chronicle of Higher Education, *September 3, 1986, with permission from the publisher.*

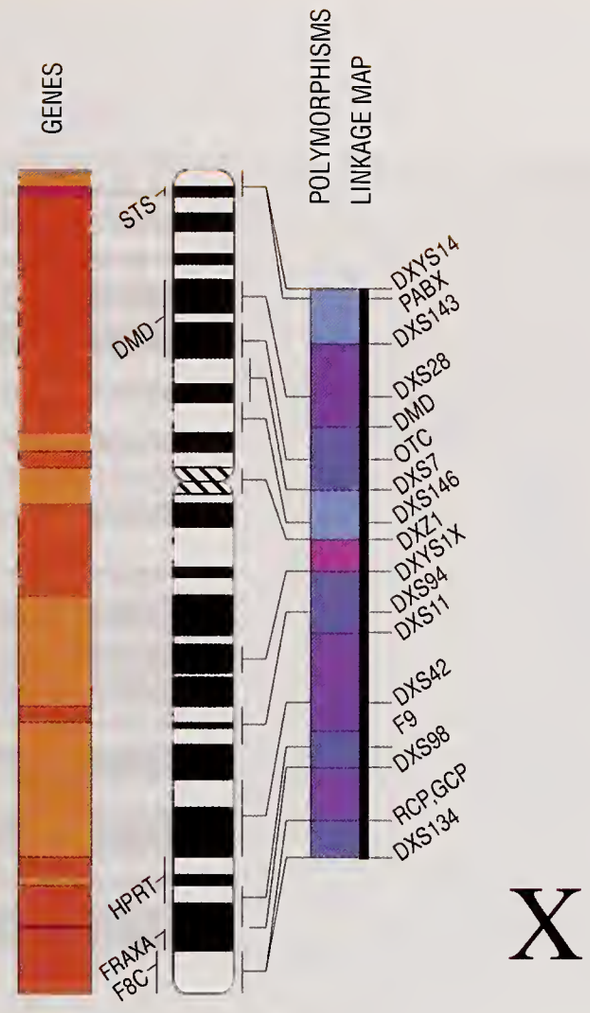


Figure 3. A schematic representation of the X chromosome showing the location of the characteristic banding patterns and the distribution of several of the genes that have been localized to this chromosome.

Excerpt from Stephens, J.C., Mador, M.L., Cavanaugh, M.L., Gradie, M.I., and Kidd, K.K. The Human Genome Map. Science *250: October 12, 1990. Copyright 1990 by the AAAS.*

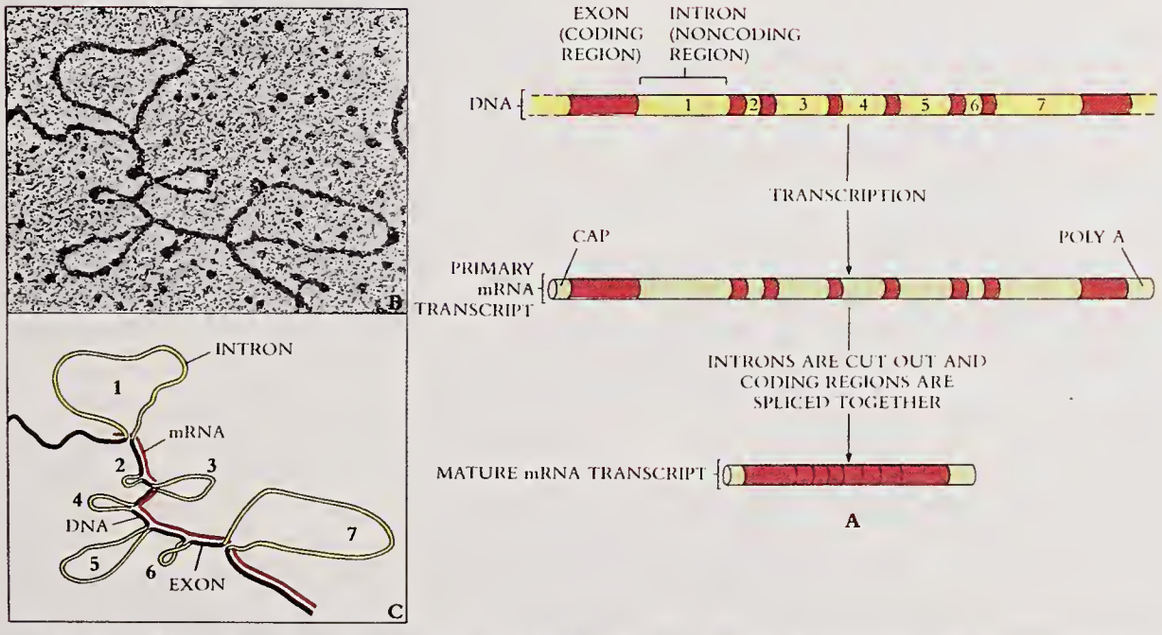


Figure 4. The primary RNA transcript of a gene is very much larger than the message that is actually translated to form a protein. A variable number of noncoding sequences called introns are edited out, and the remaining exons are spliced together by a complex mechanism shown in the drawings to the left.

From Dugaiczyk, A., Woo, S.L.C., Colbert, D.A., Lai, E.C., Mace, M.L., Jr., and O'Malley, B. 1979. Proc Natl Acad Sci USA *76:2253–2257.*

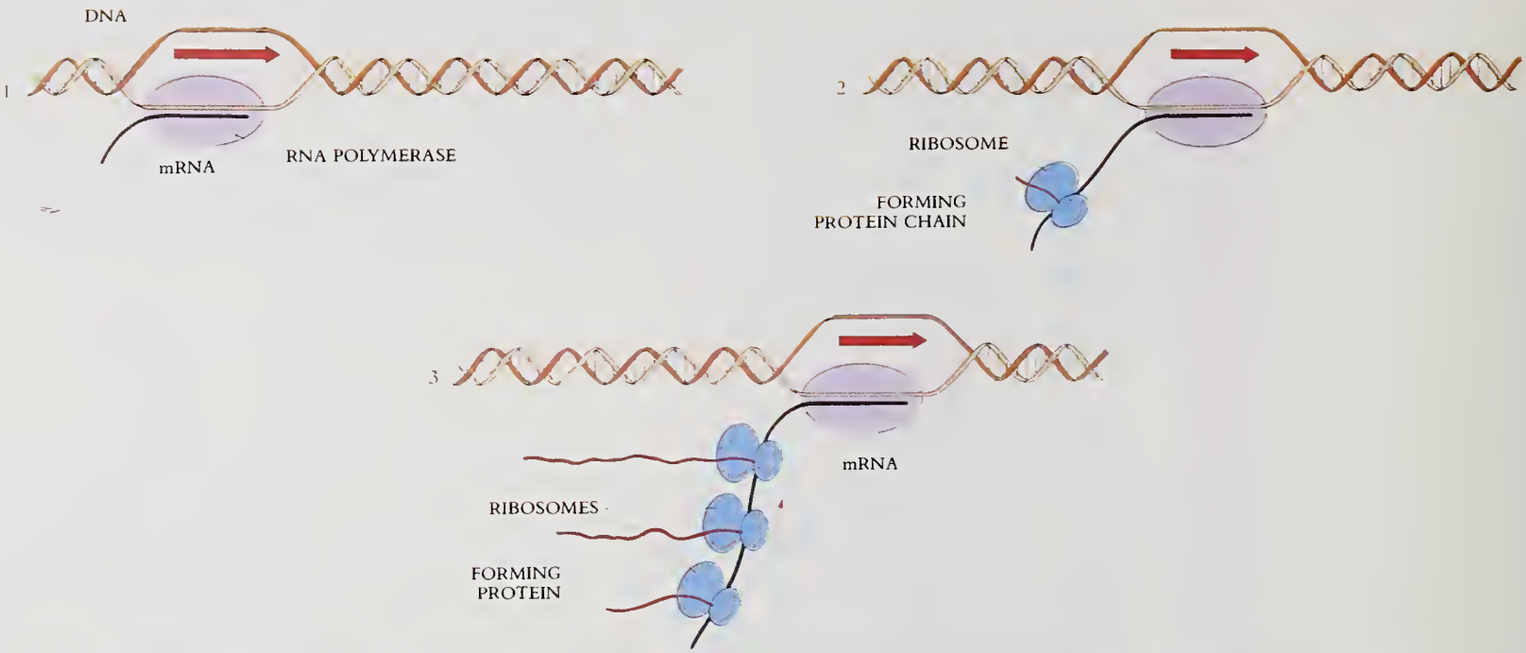


Figure 5. Transcription of a gene by the enzyme RNA polymerase and the subsequent translation of this messenger RNA (mRNA) by ribosomes to form a protein.

From Raven, P.H., and Johnson, G.B. 1988. Understanding Biology. *St. Louis, MO: Times Mirror/Mosby College Publishing, p. 267.*

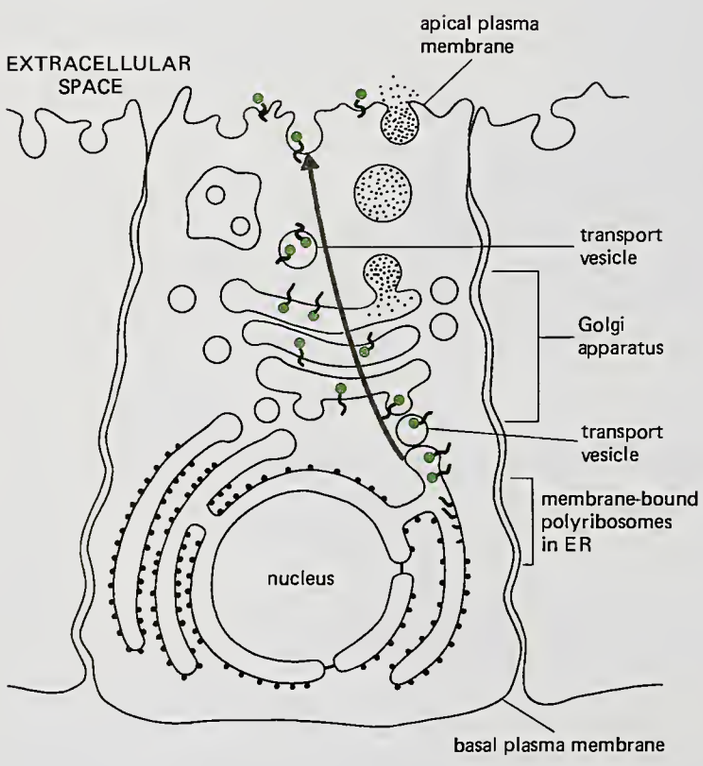


Figure 6. The synthesis and intracellular trafficking of proteins from the endoplasmic reticulum through the Golgi apparatus to the cell membrane.

From Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. 1989. Molecular Biology of the Cell, *2nd edition. New York: Garland, p. 455.*

genes it uses (and thus the kinds of proteins it produces) or the amounts of each protein it makes (Figure 5)? In the next section we describe the considerable progress that has been made recently in deciphering the DNA sequence elements that determine whether a gene is transcribed. Proteins, commonly known as **transcriptional factors**, bind to these DNA sequence elements and determine when, and how frequently, a gene is transcribed. Since this often influences how much of the encoded protein is produced by a cell, such information provides one level of insight about how cells differ from one another.

Another large issue concerns the way in which proteins are deployed in the cell. A cell is not just a bag of randomly distributed molecules; it has a highly organized internal structure. The DNA that makes up the genes and the machinery for gene transcription are packaged in the nucleus, which is surrounded by a membrane with distinctive pores that separates it from the **cytoplasm**, which comprises the rest of the cell. Nearly all cellular functions are compartmentalized in other cellular structures collectively referred to as **organelles**. Among the more prominent organelles in animal cells are the **mitochondria**, which are the cells' principal energy source; **lysosomes**, which are concerned largely with the degradation of foreign materials and of cellular proteins whose functions have been fulfilled; the **rough endoplasmic reticulum**, a complex, ribosome-studded system of membranes responsible for the synthesis of proteins secreted by the cells; and the **Golgi apparatus**, which both modifies proteins (by adding other chemical groups such as sugars) and also packages them for transport to their appropriate locations, such as the cell membrane. Most of these organelles are themselves surrounded by membranes that separate their functions from those of the rest of the cell. It is easy to see how such a compartmentalized structure allows the cell to organize its different processes efficiently, but it poses an organizational problem that is of considerable current interest in cell biology: How are particular proteins routed to the correct organelles? As we began to learn about the structures of individual proteins, it was discovered that there are specific "signals" built into proteins that target them to particular organelles or particular locations within the cell and that there are distinctive cellular machineries that "detect" these signals and steer the proteins in particular directions (Figure 6). Thus certain proteins are directed to the nucleus, while others are targeted for insertion into the surface membrane of the cell, and yet others are destined for export out of the cell as secretory products.

At another level of organization, it has become evident that the organelles within a cell are also not distributed randomly. In many cells one can identify a distinct "top" and "bottom." Other cells, while not polarized in this manner, have asymmetric structures arranged in such a way that given organelles are distributed in different, but quite reproducible, patterns. Still other cells, especially nerve cells, have unusual extensions or processes that may be many hundred times as long as the body of the cell. In each case organelles have to be transported to particular locations and maintained there; they do not drift about haphazardly inside the cell.

The asymmetric shapes of cells and the locations of their organelles both rely on cellular structures known collectively as the **cytoskeleton**. The cytoskeleton consists of several types of elongated threads or filaments (**microfilaments**, **microtubules**, **intermediate filaments**), each made of specific proteins that are so designed as to assemble spontaneously into filaments. These cytoskeletal elements serve as a form of internal scaffolding to maintain the shape of the cell, and as a system of tracks along which organelles can be transported. Recent research has disclosed a variety of proteins that function as molecular motors that can move proteins and organelles that attach to them along particular cytoskeletal filaments to various locations within the cell. We also know that the appropriate organization of cytoskeletal filaments and motor proteins can, in some cases, contribute to **cell motility**, that is, the movement of the entire cell from one location to another. Such cellular movements are especially important in development but continue to play an integral role in the life of certain cells even in adult life.

All these processes—gene transcription, protein targeting, organelle movement, and cell motility—must be carefully regulated so that cells respond appropriately to different situations. The same is true of many other cellular processes. For example, the proliferation of cells that takes place by **cell division** involves copying or replicating the genes, the breakdown of the nuclear membrane, the separation of the duplicated chromosomes into two equivalent sets, division of the cell into two daughter cells, the re-

formation of a nucleus in each of the daughter cells, and finally the resumption of normal function in both cells. This whole process, which is known as the **cell cycle**, takes place whenever cells divide and remains an important part of the life of all but a few cell types. Many types of cells —for example, the cells in the blood and skin— are continually being formed and replaced. Other cells proliferate rarely, and some divide only during early development; the nerve cells that make up the brain, for example, proliferate rapidly during development, but no further cell division occurs before death of the individual some 70 or more years later. Obviously, cell proliferation must be tightly controlled. Recent advances have shown that the cell cycle is controlled by a set of proteins whose role is to modify other proteins selectively and thus regulate their functions. One common way in which proteins are modified (but certainly not the only one) is to attach a small chemical group, such as a phosphate group, to a protein. Proteins that attach phosphate groups to other proteins are called **protein kinases**. Control of many aspects of the cell cycle and, indeed, of many other cellular functions relies on complex control networks of protein kinases acting on key proteins at pivotal stages in the life of the cell.

Thus far we have considered only processes occurring within a cell. An important related set of issues concerns how cells interact with each other and how they respond to the external environment. Each cell is surrounded by a **surface** (or **plasma**) **membrane**, which serves as a selective barrier separating the inside of the cell from the world outside itself. Embedded in this membrane are several types of proteins. One essential class are **transporters**, specialized for the ordered movement in and out of the cell of nutrients, ions and other small molecules that are essential for normal cell function.

A second group of cell surface proteins are **receptors**, which bind other types of molecules that interact with the cell. As the name suggests, receptors serve to receive input from the cell's external environment. They are of many different types. The largest group binds peptide hormones or diffusible factors produced locally or at a distance by other cells, but another important group serves to transport materials like cholesterol from outside to the interior of the cell. Typically these receptors have three parts: an external part or ligand-binding domain that can bind the hormone or diffusible factor, a transmembrane domain that spans the cell membrane, and an intracellular part that can interact with internal components of the cell. The binding of a hormone or other diffusible factor to such a receptor triggers in some way, as yet undetermined, a signal inside the cell. These signals are of many types. Some receptors are protein kinases that are selectively activated by binding the appropriate external factor; others, when activated, lead to the release of diffusible, small molecules, such as calcium ions or cyclic nucleotides. These diffusible **second messenger** molecules in turn activate other control mechanisms inside the cell, including protein kinases and other regulatory molecules. In this way the triggering of a receptor from outside cells can result in a cascade of events that ultimately controls the various intracellular processes discussed earlier. One of the current "hot topics" in cell biology research concerns the nature and mechanisms of cell surface receptor signaling and the control circuits inside cells that link receptor activity to other control mechanisms, including those that regulate gene function and the control of the cell cycle.

A special subset of this class of receptors are those that respond to the release of chemical signals (**transmitters**) at the specialized endings of nerve cell processes (Figure 7). The released neurotransmitters bind to the external part of the cell surface receptor and in doing so may open an ion channel or trigger the activation of a second intracellular message. Since the majority of nerve signals are transmitted from cell to cell in this way, the analysis of this class of receptors is (as we shall see in the section on neuroscience) one of the central issues in contemporary neuroscience.

Another class of cell surface receptors is involved in the adhesion of cells, either to their neighbors or to the **extracellular matrix**, a complex group of secreted proteins and polysaccharides that assemble into an organized meshwork on the cell surface. Depending upon the cell type and environment, the extracellular matrix performs various functions (Figure 8). In a petri dish, for example, the extracellular matrix provides a cushion on which the cell sits. In the epidermis, the extracellular matrix helps to form the basement membrane, which anchors the epidermis to the rest of the skin. In connective tissues, the extracellular matrix completely surrounds most cells and is often more extensive in its distribution than the cells themselves. In

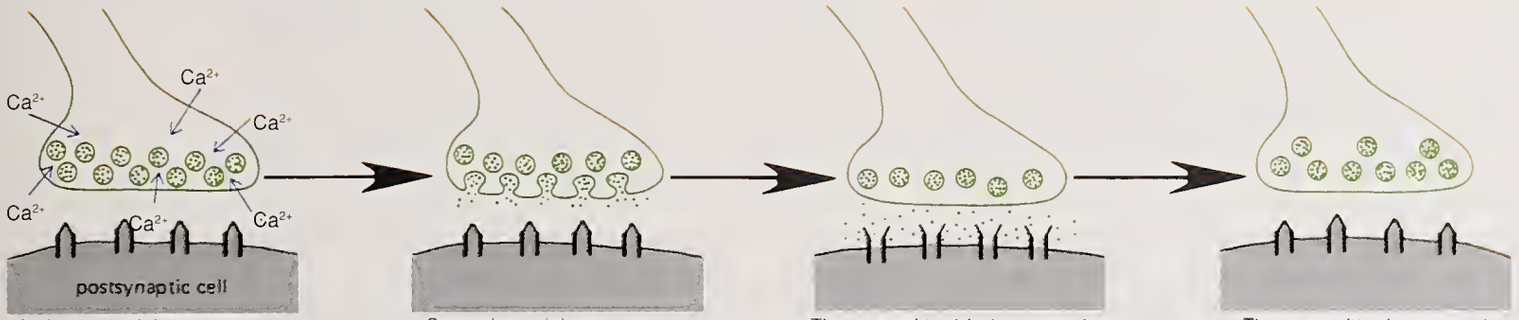


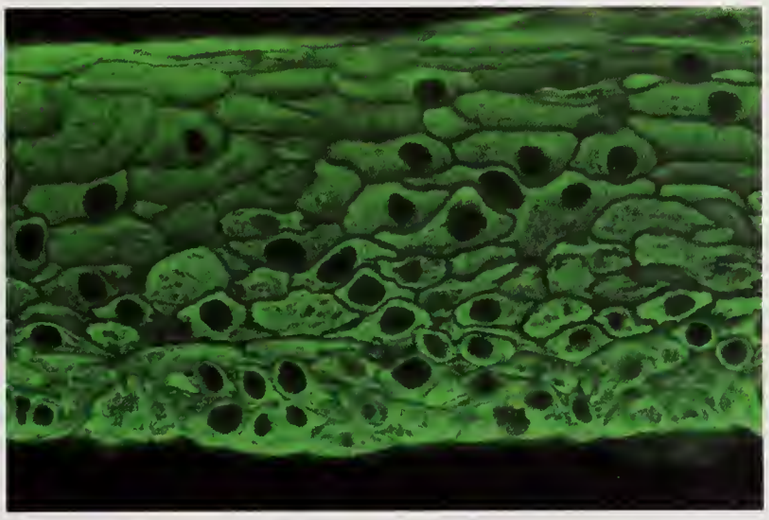
Figure 7. A summary of the essential events at a chemical synapse following the arrival of an action potential in the axon terminal.

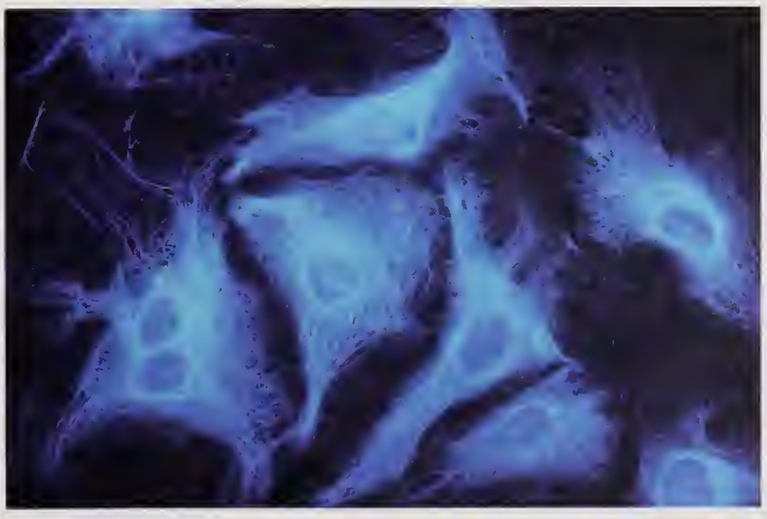
Adapted from Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. 1989. Molecular Biology of the Cell, *2nd edition. New York: Garland, p. 1077.*

Figure 8. A: A section of human skin stained with an antibody against the protein keratin. Note how the cells are packed together to form the distinctive structure of skin. The dark structures within each cell are the nuclei that do not contain keratin and so are not labeled.

B: A fluorescence micrograph showing the presence of keratin filaments in cultured skin cells. The keratin filaments have been revealed by a fluorescently labeled antibody against the keratin protein.

Courtesy of Elaine V. Fuchs.


A.


B.

this case, the extracellular matrix helps to provide the body's architectural framework.

Cellular adhesion, which plays a crucial role in cell, tissue, and organ structure and in cell movements, depends on specialized cell adhesion receptors that are connected to the intracellular cytoskeleton. It is also likely that cells can signal to one another via cell adhesion receptors. Decisions as to whether a cell remains stationary, or where and when it should move, and what shape it should take up, are all affected by cell adhesion events. And these, in turn, are largely dependent on the pattern and functions of cell adhesion receptors deployed on the surfaces of the cells.

Elucidating the normal functions of cells is an important first step in understanding how these processes go awry in a number of human diseases. For example, it is now known that alterations in certain normal genes (called **oncogenes**) can contribute to cancer. It is also clear that many oncogenes encode proteins that are involved in the regulation of specific cellular functions: some oncogene-encoded proteins are growth factors; others are cell surface receptors; yet others are signaling molecules, protein kinases, and transcription factors. Many other human diseases are known to be due to defects of one or another of the cellular processes reviewed above. Thus disturbances in insulin production lead to diabetes mellitus, defects in the extracellular matrix can cause osteogenesis imperfecta, and abnormalities in cell adhesion receptors may result in various bleeding disorders. Indeed, one of the major insights in pathology and medicine is that all disease processes are ultimately attributable to the breakdown of one or more cellular functions. In the twenty-first century, we can be fairly certain that medicine will be concerned largely with the identification and treatment of specific disordered cell functions. We already know of many disorders that can be attributed to disturbances in particular organelles. What is particularly encouraging to researchers in cell biology today is that new avenues are beginning to be perceived for therapy, as the molecular bases of various disordered functions become known. Indeed, one of the especially appealing aspects of modern cell biological research is the immediacy with which fundamental research advances are having an impact on medically important problems.

Investigators in the Cell Biology and Regulation Program

Alexander-Bridges, Maria C., M.D., Ph.D.
Beach, David H., Ph.D.
Bennett, G. Vann, M.D., Ph.D.
Beutler, Bruce A., M.D.
Bevilacqua, Michael P., M.D., Ph.D.
Blackshear, Perry J., M.D., D.Phil.
Blobel, Günter, M.D., Ph.D.
Bonadio, Jeffrey F., M.D.
Brugge, Joan S., Ph.D.
Campbell, Kevin P., Ph.D.
*Caron, Marc G., Ph.D.**
Carroll, Sean B., Ph.D.
Chin, William W., M.D.
Crabtree, Gerald R., M.D.
Craig, Nancy L., Ph.D.
Cunningham, James M., M.D.
Davis, Laura I., Ph.D.
Davis, Roger J., M.D.
Dreyfuss, Gideon, Ph.D.
Emr, Scott D., Ph.D.
Esmon, Charles T., Ph.D.
Exton, John H., M.D., Ph.D.
Fuchs, Elaine, Ph.D.
Ganem, Donald E., M.D.
Garbers, David L., Ph.D.
Gething, Mary-Jane H., Ph.D.
Glomset, John A., M.D.
Gomer, Richard H., Ph.D.
Grosschedl, Rudolf, Ph.D.
Habener, Joel F., M.D.
Heintz, Nathaniel, Ph.D.
Hynes, Richard O., Ph.D.
Isberg, Ralph R., Ph.D.
Kaback, H. Ronald, M.D.
Kim, Peter S., Ph.D.
Kirkegaard, Karla A., Ph.D.
Kobilka, Brian K., M.D.
Lai, Michael M.-C., M.D., Ph.D.
Lamb, Robert A., Ph.D., Sc.D.
Lefkowitz, Robert J., M.D.
Lehmann, Ruth, Ph.D.
Maas, Richard L., M.D., Ph.D.
Maller, James L., Ph.D.
Massagué, Joan, Ph.D.
McKnight, Steven Lanier, Ph.D.
Nusse, Roel, Ph.D.
O'Donnell, Michael E., Ph.D.
Parker, Keith L., M.D., Ph.D.
Pike, Linda J., Ph.D.
Sadler, J. Evan, M.D., Ph.D.
Schekman, Randy W., Ph.D.
Schoolnik, Gary K., M.D.
Shenk, Thomas E., Ph.D.
Sherr, Charles J., M.D., Ph.D.

Simpson, Larry, Ph.D.
Spradling, Allan C., Ph.D.
Steiner, Donald F., M.D.
Welsh, Michael J., M.D.
Williams, Lewis T., M.D., Ph.D.
Wilson, James M., M.D., Ph.D.
Wolin, Sandra L., M.D., Ph.D.
Yanagisawa, Masashi, M.D., Ph.D.

* *This investigator was appointed after manuscripts were submitted for publication. His research will be described in the next volume.*

Genetics Program

The emergence of what is sometimes referred to as the "new genetics" has contributed more to our fundamental understanding of biology and medicine in the past two decades than any other advance in biomedical science. Not surprisingly, this development has come to assume a central place in all biological research, making it possible to examine biological processes at a level of resolution that was considered quite impossible just 20 years ago. "Gene cloning," "recombinant DNA technology," "genetic engineering," and the "Human Genome Initiative" are phrases that have entered everyday language, but the possibilities they offer for major advances in biology and medicine have yet to be fully appreciated. Given the central role of genetics in modern biology and the current sense of excitement that the new genetics has generated, it is appropriate that the Genetics Program should be the largest research program within HHMI.

Historically, one of the first applications of Mendel's classic laws of inheritance was to the analysis of certain human diseases. In the early years of this century, Archibald Garrod, an English physician, noted that a number of relatively rare diseases tended to occur in families, often in families with consanguineous marriages. The pattern of occurrence of these rare diseases followed Mendel's laws, discovered almost 50 years earlier. Mendel's work on inherited characteristics in plants was neglected for many years, but after it was rediscovered, around the turn of the century, it was quickly established that the laws of genetics are universal and govern inheritance in organisms as disparate as peas and worms, mice and fruit flies, bacteria and human beings. Beginning with Garrod, these genetic laws were applied to a host of inherited diseases, as it was realized that the genetic makeup of an individual can have a profound effect on his or her health and well-being. However, an understanding of what genes are and how they function had to await the discovery, in 1944, that the genetic material is DNA, and, in 1953 and 1961, of the double helix and the genetic code, respectively.

Much of our early understanding of the action of genes came from experiments that took advantage of the universality of gene action by using simple organisms, especially bacteria and their viruses, as model systems. The cardinal discovery was the identification of deoxyribonucleic acid (DNA) as the fundamental chemical in which genetic information is encoded. But the finding that DNA has a double-stranded, mirror image structure provided the first clear insight as to how this information could be replicated and passed on from one generation of organisms to the next (Figures 9 and 10). Understanding how the chemical language of DNA could be used to direct the synthesis of other cellular constituents, especially proteins, came with the discovery of the nature of the genetic code. These great advances will always be viewed as the high watermark of the early molecular stage of modern genetics.

Notwithstanding these dramatic developments —arguably the most important in biology since the publication of Darwin's essay *On the Origin of Species* in 1859—the molecular details of the genes of higher organisms remained hidden from view by the enormous complexity of their genomes (Figure 11). Fortunately, the genes of simple organisms were accessible, because they are relatively few in number (involving in many cases an assemblage of as few as 3,000 **base pairs**, as the building blocks of DNA are called) compared to the human genome, which probably contains about 3 billion base pairs.

The problem of genetic complexity has been finally overcome in the past 10 years by the development of the powerful new genetic methods known collectively as **recombinant DNA technology**. This technology allows researchers to isolate specific genes from complex mixtures, to prepare them in sufficiently large amounts that their entire molecular structure can be determined, and to move them from one group of cells or from one organism to another, so that their functional properties can be identified and their products produced in abundance. In a number of cases, medically valuable products such as human insulin, growth hormone, antihemophilic factor, erythropoietin, and TPA (tissue

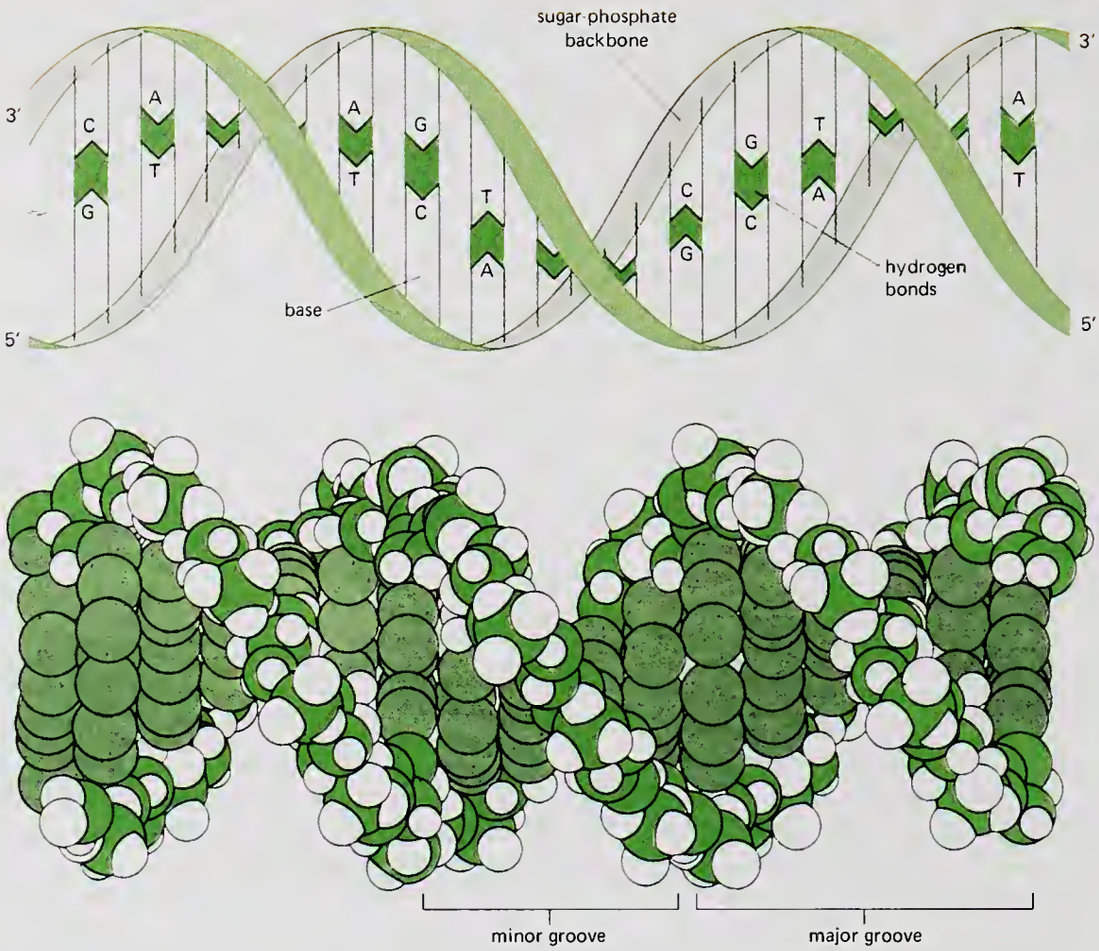


Figure 9. The famous DNA double helix.

From Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. 1989. Molecular Biology of the Cell, *2nd edition. New York: Garland, p. 99.*

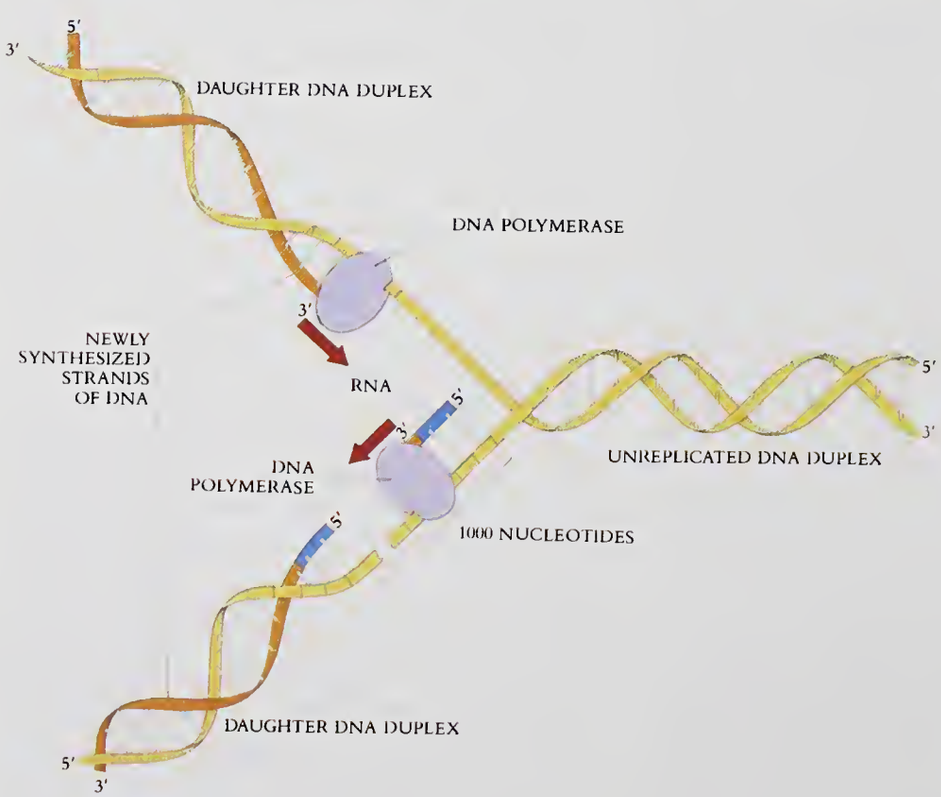


Figure 10. An illustration of how DNA is replicated so as to form two new strands that are exactly complementary to the original DNA sequences from which they derive their sequences.

From Raven, P.H., and Johnson, G.B. 1988. Understanding Biology. *St. Louis, MO: Times Mirror/Mosby College Publishing, p. 277.*

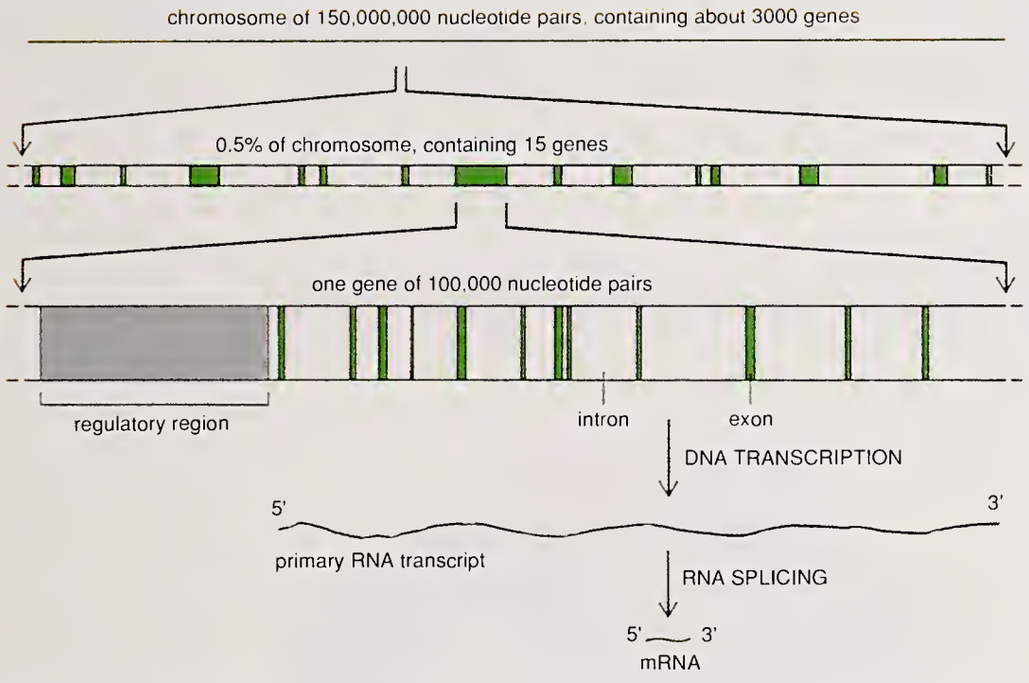


Figure 11. The organization of genes on a typical vertebrate chromosome. Proteins that bind to the DNA in regulatory regions determine whether a gene is transcribed; although often located on the 5' side of a gene, as shown here, regulatory regions can also be located in introns, in exons, or on the 3' side of a gene. The intron sequences are removed from the primary RNA transcripts that encode protein molecules to produce a messenger RNA (mRNA) molecule. The figure given here for the number of genes per chromosome is only a minimal estimate.

From Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. 1989. Molecular Biology of the Cell, *2nd edition. New York: Garland, p. 487.*

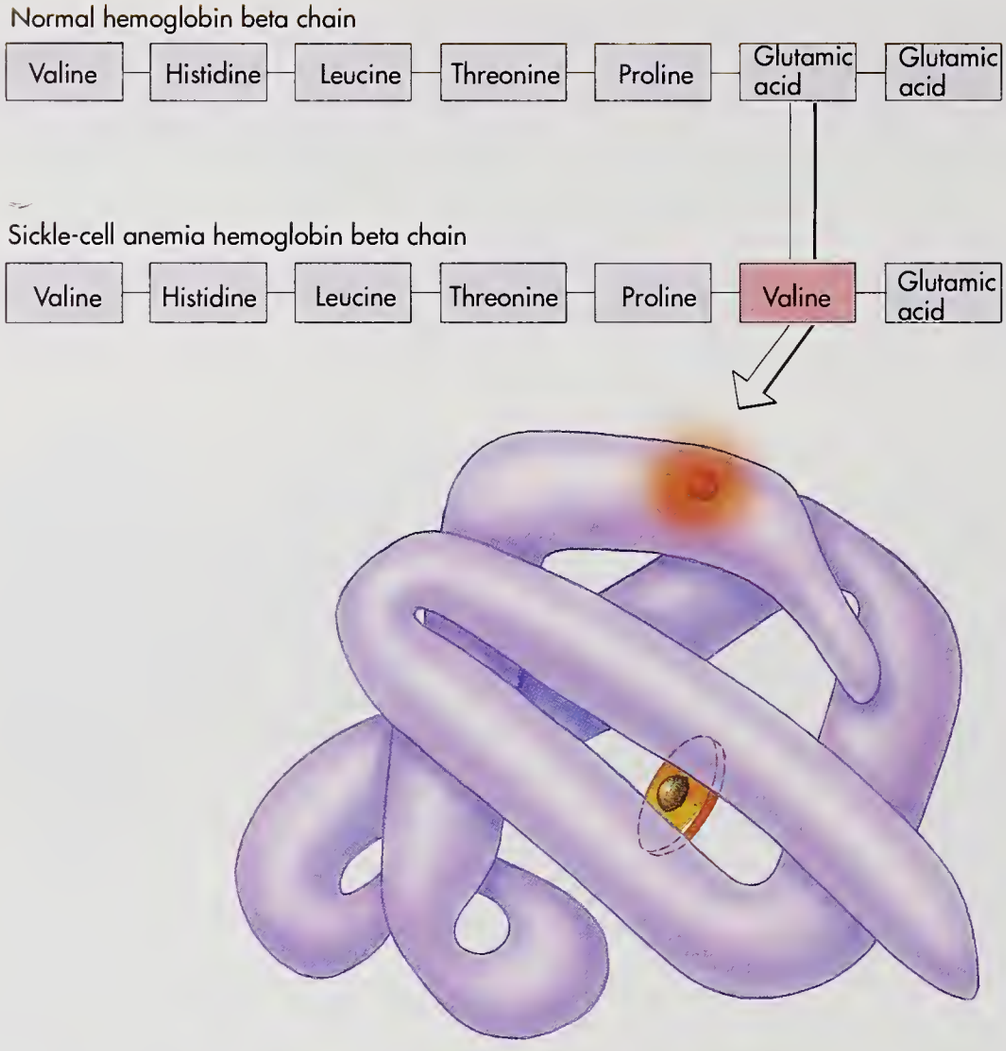


Figure 12. A drawing of the hemoglobin molecule and the single-amino acid change in the β-chain of the molecule that causes sickle cell anemia.

From Raven, P.H., and Johnson, G.B. 1988. Understanding Biology. *St. Louis, MO: Times Mirror/Mosby College Publishing, p. 255.*

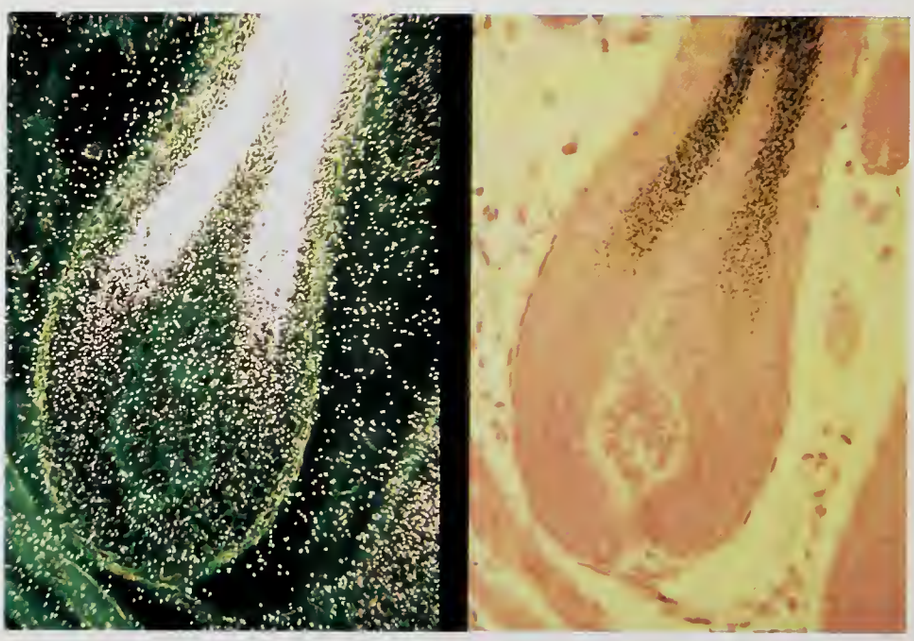


Figure 13. We can determine where particular genes are being expressed by using a radioactively labeled probe that selectively binds to the RNA transcript of the gene and examining the distribution of the radioactivity in a tissue autograph. This technique, known as in vitro *hybridization, was used in this case to show the expression of one of the keratin genes in a hair follicle. The micrograph on the left is viewed under darkfield optics (the exposed silver grains in the autoradiograph appear as white dots); that on the right is seen with conventional optics, and the radioactivity is marked by black dots.*

Courtesy of Elaine V. Fuchs.

plasminogen activator) have been commercially manufactured in this way and are now being used therapeutically.

Most of these accomplishments depend on the technique commonly referred to as **gene cloning**. In this process a gene (for example, the human insulin gene, which is a few thousand base pairs in length) is isolated from a complex mixture of human DNA fragments and transferred into the genetic apparatus of a much simpler organism such as a bacterium. Since bacteria multiply very rapidly, the inserted human insulin gene is amplified (cloned) along with the genes of its simple host. It is then a relatively straightforward process to purify the cloned gene and to determine the sequence of nucleotides that encodes the relevant genetic instructions.

One of the most important developments to emerge from gene cloning and **sequencing** is that they permit geneticists to determine exactly how a gene may have been altered (genetic alterations of this kind are called **mutations**) to produce an inherited disease. While a majority of the investigators in the HHMI Genetics Program are directing their efforts toward understanding the principles that govern the action of genes, a large and growing number are trying to identify the mutations responsible for some of the estimated 4,000 human genetic defects. Among the genes being studied are those responsible for muscular dystrophy, cystic fibrosis, several forms of hemoglobinopathy, chronic granulomatous disease, phenylketonuria, polyposis coli, neurofibromatosis, osteogenesis imperfecta, and hemophilia. Most of the genes responsible for these diseases have already been cloned and their structures determined; in this way the precise molecular effects of the mutations that cause the disorders are now being determined. Information that has been derived in this way is being used in several ways. For example, in some cases it is being used to counsel affected families about the risks they face in having additional children; in others it is being used to develop tests that are critical for prenatal diagnosis, and, in the case of hemophilia and chronic granulomatous disease, it is already being used to provide appropriate therapy. Information about some of these genes and their protein products has also told us a great deal about the role of the relevant proteins in development (e.g., the protein **dystrophin**, which is normally found in muscle but is missing in most cases of muscular dystrophy) or about the mechanisms required for normal function (e.g., in the lungs and pancreas in cystic fibrosis, or for the mechanisms of blood clotting in hemophilia). It is the continuous interplay between the development of new basic knowledge and its application to the understanding of human disease that is so dramatically informing modern medicine.

The execution of a genetic program in a cell obviously cannot be left to chance. To bring about an ordered series of changes in a developmental or metabolic process, each genetic instruction must be activated at a specific time, and the product of the gene must be produced in an amount commensurate with the needs of the organism. As was pointed out in the section on cell biology, different sets of genes are expressed in each cell type so, for example, many of those expressed in muscle cells are likely to be quite different from those expressed in skin cells or in bone. Since with few exceptions (one of which will be discussed in the section on immunology) each cell contains the same genetic information, there must be some mechanism or mechanisms that insure that the appropriate genes are being expressed in each cell type. The processes that govern these crucial steps are generally referred to as **gene regulation**.

Once again, the study of inherited human diseases has revealed mutations that can disrupt those regulatory programs. For example, it is now known that certain inherited anemias are brought about by specific mutations that affect the regulatory apparatus of the genes that encode the red blood cell protein, hemoglobin (Figure 12). An even larger number of examples could be cited from work on simpler organisms to show how powerful this mutational analysis has been for our understanding of genetic regulation. However, it will suffice to say here that the work of a considerable number of HHMI investigators has this as its primary theme.

As discussed in the section on cell biology and regulation, genetic information is usually conveyed through a series of steps from the nucleus of the cell to the surrounding cytoplasm, where it is decoded or translated to form the protein products that execute the relevant genetic program. This process involves making many copies of the gene, in the closely related chemical, RNA. The process of copying a gene from DNA into RNA is called **transcription**. This is one of the key points at which the flow of genetic information can be regulated. Genes that are not scheduled for expression in a particular cell type are not

transcribed: while still present, they remain silent and ineffective.

Since it plays such a central role in regulating the expression of genetic information, it is not surprising that the study of transcription is one of the most active areas in modern genetic research (Figure 13). Investigators have taken two successful approaches in an effort to understand this process. One involves the study of mutations that disrupt regulation in a variety of organisms. The other is biochemical and involves isolating specific proteins that are parts of the transcription machinery and then determining their mode of action at a molecular level. The ultimate goals of both approaches are to provide a complete account, in chemical terms, of the processing of genetic information and to provide a sound basis for our understanding of the types of disturbances in this process that can lead to various genetic defects and many diseases in adult life.

It is perhaps worth commenting on the broad spectrum of organisms that are currently used as models in genetic investigations. Yeast, worms, fruit flies, bacteria, viruses, mice, and human beings all provide instructive paradigms. For example, because the structure of the nervous system is relatively well understood in the primitive worm, *Caenorhabditis elegans*, it has proved to be particularly useful for elucidating the genetic program that directs the formation of the nervous system. The fruit fly *Drosophila melanogaster*, which has an uncommonly rich background of genetic variation, has proved to be the most valuable organism for studying the genes involved in early embryonic development. Similarly, the mouse, with a generation time of just a few months, is a convenient stand-in for genetic experiments that cannot be contemplated in most mammals or human beings. Thus it has become an everyday technique to introduce new genes into the genetic makeup of mice. Many of the resulting animals, called **transgenic mice**, have proved to be powerful models for certain human diseases. Transgenic mice that are genetically cancer-prone or that have sickle cell anemia have been generated recently and are proving useful not only for understanding the molecular bases of these disorders but also for exploring possible means for their treatment. Complementing this approach has been the development of techniques that allow one to knock out specific genes selectively and to generate mice of essentially any specified genotype. The power of this new methodology (that involves **homologous recombination**) for understanding mammalian development, for studying the function of the nervous system, and for modeling known human diseases has already captured the attention of many geneticists, including several in the HHMI Genetics Program. At the other end of the spectrum, the simple baker's yeast, *Saccharomyces cerevisiae*, is proving to be especially useful for cloning large segments of mammalian chromosomes. Special carrier elements called YACs (yeast artificial chromosomes) have been developed that can accommodate up to 500,000 base pairs of DNA. Since this is more than 20 times as large as the fragments that have traditionally been cloned in bacteria and viruses, this approach and the associated method for separating large DNA fragments are beginning to play an important part in the international effort to map the human genome.

The development of a complete **map of the human genome** is one of the great challenges engaging the attention of geneticists worldwide, including several HHMI investigators. Such a map will permit the identification of genes that are close to, or responsible for, a large number of as yet uncharacterized genetic diseases. Furthermore, knowledge of the map will allow the development of easily identifiable genetic markers for specific diseases. These genetic markers, called **restriction fragment length polymorphisms (RFLPs)**, can be used to detect carriers of many genetic disorders, to determine paternity, and to identify individuals for forensic purposes (a process commonly referred to in the press as **DNA fingerprinting**).

Four or five years ago an extremely sensitive technique for detecting these DNA marker fragments was developed. This technique (the **polymerase chain reaction [PCR]**) greatly amplifies the genetic signal and has made these studies both simpler and faster. With this and other emerging technologies, the structures of the genomes of several simple bacteria should be known in the near future, and the way is already clear to begin the systematic structural analysis of the genomes of a number of more complex organisms, such as the nematode worm *C. elegans* and the fruit fly *Drosophila*. It is confidently predicted that with further improvements in the technology for sequencing genes, the complete structure of the human genome should be known within 12 or 15 years.

While the new genetics has opened the door to an understanding of an important range of fundamental problems such as the mechanism of chro-

mosome replication, the genetic basis of embryonic development, and the control of gene regulation, it also holds promise for an assault on the even more complex problems that confront those interested in human disease. Cancer has a clear genetic basis. Specific genes, generally those concerned with cell growth and gene regulation, have been shown to be involved in many different malignancies. Other disorders, such as heart disease and hypertension, have clear genetic components. Even distressing behavioral disorders like manic depression and schizophrenia have a genetic basis that may offer important clues to their nature and perhaps, in time, to their treatment.

The real power of the new genetics is that it allows investigators to approach these complex problems using a reductionist approach. If, for instance, a gene contributes to the development of a behavioral disorder, it is reasonable to assume that it also plays an important role in programming normal behavior. By identifying such genes and understanding the chemistry of their protein products, we should obtain important insights about the biological basis of behavior. Thus the ultimate promise of the new genetics is that it will help us to understand diseases that are not commonly thought of as being primarily genetic and, in doing so, it will tell us much about those biological processes that lie at the very core of our humanity.

Investigators in the Genetics Program

Barsh, Gregory S., M.D., Ph.D.
Beachy, Philip A., Ph.D.
Beaudet, Arthur L., M.D.
Bell, Graeme I., Ph.D.
Bellen, Hugo J., D.V.M., Ph.D.
Belmont, John W., M.D., Ph.D.
Brown, Patrick O., M.D., Ph.D.
Capecchi, Mario R., Ph.D.
Caskey, C. Thomas, M.D.
Cech, Thomas R., Ph.D.
Church, George M., Ph.D.
Cohen, Stephen M., Ph.D.
Collins, Francis S., M.D., Ph.D.
Corden, Jeffry L., Ph.D.
Cullen, Bryan R., Ph.D.
Desiderio, Stephen V., M.D., Ph.D.
Desplan, Claude, Ph.D.
Donelson, John E., Ph.D.
Duyk, Geoffrey M., M.D., Ph.D.
Feinberg, Andrew P., M.D., M.P.H.
Francke, Uta, M.D.
Friedman, Jeffrey M., M.D., Ph.D.
Geliebter, Jan, Ph.D.
Gesteland, Raymond F., Ph.D.
Ginsburg, David, M.D.
Gitschier, Jane M., Ph.D.
Henikoff, Steven, Ph.D.
Horwich, Arthur L., M.D.
Kadesch, Thomas R., Ph.D.
Kan, Yuet Wai, M.D., D.Sc.
Kaufman, Thomas C., Ph.D.
Kunkel, Louis M., Ph.D.
Kurnit, David M., M.D., Ph.D.
Laimins, Laimonis A., Ph.D.
Lalouel, Jean-Marc, M.D., D.Sc.
Leder, Philip, M.D.
Ledley, Fred D., M.D.
Liebhaber, Stephen A., M.D.
Lindquist, Susan L., Ph.D.
Lowe, John B., M.D.
Malim, Michael H., Ph.D.[*]
Nabel, Gary J., M.D. Ph.D.
Nadal-Ginard, Bernardo, M.D., Ph.D.
Nathans, Daniel, M.D.
Nathans, Jeremy, M.D., Ph.D.
Nevins, Joseph R., Ph.D.
Nussbaum, Robert L., M.D.
Olson, Maynard V., Ph.D.
Orkin, Stuart H., M.D.
Overbeek, Paul A., Ph.D.
Page, David C., M.D.
Palmiter, Richard D., Ph.D.
Perrimon, Norbert, Ph.D.
Reeders, Stephen T., M.D.
Rosbash, Michael, Ph.D.
Sakonju, Shigeru, Ph.D.
Seidman, Jonathan G., Ph.D.
Singh, Harinder, Ph.D.
Soriano, Philippe M., Ph.D., D.Sc.
Steitz, Joan A., Ph.D.
Sternberg, Paul W., Ph.D.
Sukhatme, Vikas P.. M.D., Ph.D.
Taub, Rebecca A., M.D.
Thummel, Carl S., Ph.D.
Tilghman, Shirley M., Ph.D.
Tjian, Robert, Ph.D.
Valle, David L., M.D.
Warren, Stephen T., Ph.D.
Weintraub, Harold M., M.D., Ph.D.
White, Raymond L., Ph.D.
Williams, David A., M.D.
Woo, Savio L. C., Ph.D.
Young, Michael W., Ph.D.

[*] *This investigator was appointed after manuscripts were submitted for publication. His research will be described in the next volume.*

Immunology Program

One of the most important developments that occurred during the evolution of vertebrates was the emergence of protective mechanisms that enable animals to defend themselves against invasion by foreign organisms, especially viruses, bacteria, and a number of disease-causing parasites. These defenses employ two distinct but related strategies: the first detects, with exquisite sensitivity, the presence of foreign organisms, cells, or proteins; the second involves a series of mechanisms that act cooperatively to rid the body of the invading organisms or, at the least, to neutralize their potentially harmful effects. The task of effecting both strategies falls to the **immune system**.

Recognizing the strategic importance of the immune system in both health and disease, the Institute selected Immunology to be one of its earliest research programs. The wisdom of that decision has been amply borne out by the truly remarkable progress that has been made in immunology in the past two decades. With the notable exception of molecular genetics, no field of biomedical research has witnessed such an astonishing series of successes at almost every level, from understanding the immune system's unique recognition mechanisms to the elucidation of the cellular and chemical means used to destroy or neutralize invading organisms.

The body's initial line of defense against invasion by foreign organisms is the continuously patrolling system of **macrophages** and other types of blood-borne phagocytic cells that act both as an early warning system and as a "first-strike" defense. These cells respond by ingesting and breaking up the invading organisms and by releasing soluble signaling molecules like **interleukin-1** that serve, among other things, to mobilize the next line of defense, the **immune response** (Figure 14). This response involves two classes of lymphocytes, called T and B cells, reflecting their origin from the thymus and bone marrow, respectively.

The first step in the immune response is the activation of a special subclass of **T lymphocytes** called **helper T cells**. Macrophages present fragments of foreign proteins, or **antigens**, on their surfaces. Recognition of these antigens by specialized receptors found on helper T cells then initiates the two responses: a **cell-mediated immune response** and a **humoral immune response**. The cell-mediated response involves principally the stimulation of another subclass of T lymphocytes called **cytotoxic T cells** that recognize and destroy infected cells. The humoral response, on the other hand, involves the activation of the second major class of lymphocytes, the **B cells**, to produce circulating **antibodies**. Antibodies recognize and neutralize soluble antigens and mark cells or viruses bearing antigens for destruction by phagocytic cells.

One of the central problems in immunology concerns the way in which lymphocytes recognize antigens. The complexity of this problem may be gauged from the observation that humans and other higher vertebrates are capable of forming antibodies against virtually any molecule or part of a molecule (**epitope**), including even those that do not occur naturally but are chemically synthesized in a laboratory. How does this occur? And how does the immune system distinguish foreign molecules from those produced by its own cells? In a word, how do lymphocytes distinguish **self** from **non-self**?

The key to the first issue, as we now know, is to be found in the almost unlimited variety of receptors on the surfaces of lymphocytes. The discovery of how just a few hundred genes are capable of producing such extraordinary receptor diversity is one of the great success stories of modern immunology. The essential features of the immune system's capacity for generating molecular diversity can be summarized briefly by stating that **lymphocyte receptors** are formed by pairs of protein chains that are chemically linked to form a complex receptor structure. Each chain of the pair has a constant domain and a variable domain. The **variable** domain of the two chains is responsible for antigen recognition and the discrimination between self and non-self. The constant (**invariant**) domain is physically linked to other membrane proteins of the receptor complex that activates the lymphocyte's internal signaling and effector mechanisms. T and B cells triggered via their antigen receptors respond to auxiliary signaling molecules by proliferating and differentiating to a mature effector stage. In the case of B cells, the maturation process ultimately results in the generation of **plasma cells** that produce large amounts of antibody for secretion into bodily fluids, chiefly the bloodstream.

The complex structure of the variable parts of the receptors is due to several processes. First, and most important, the genes responsible for this portion of the receptor are assembled from a large number of different gene segments

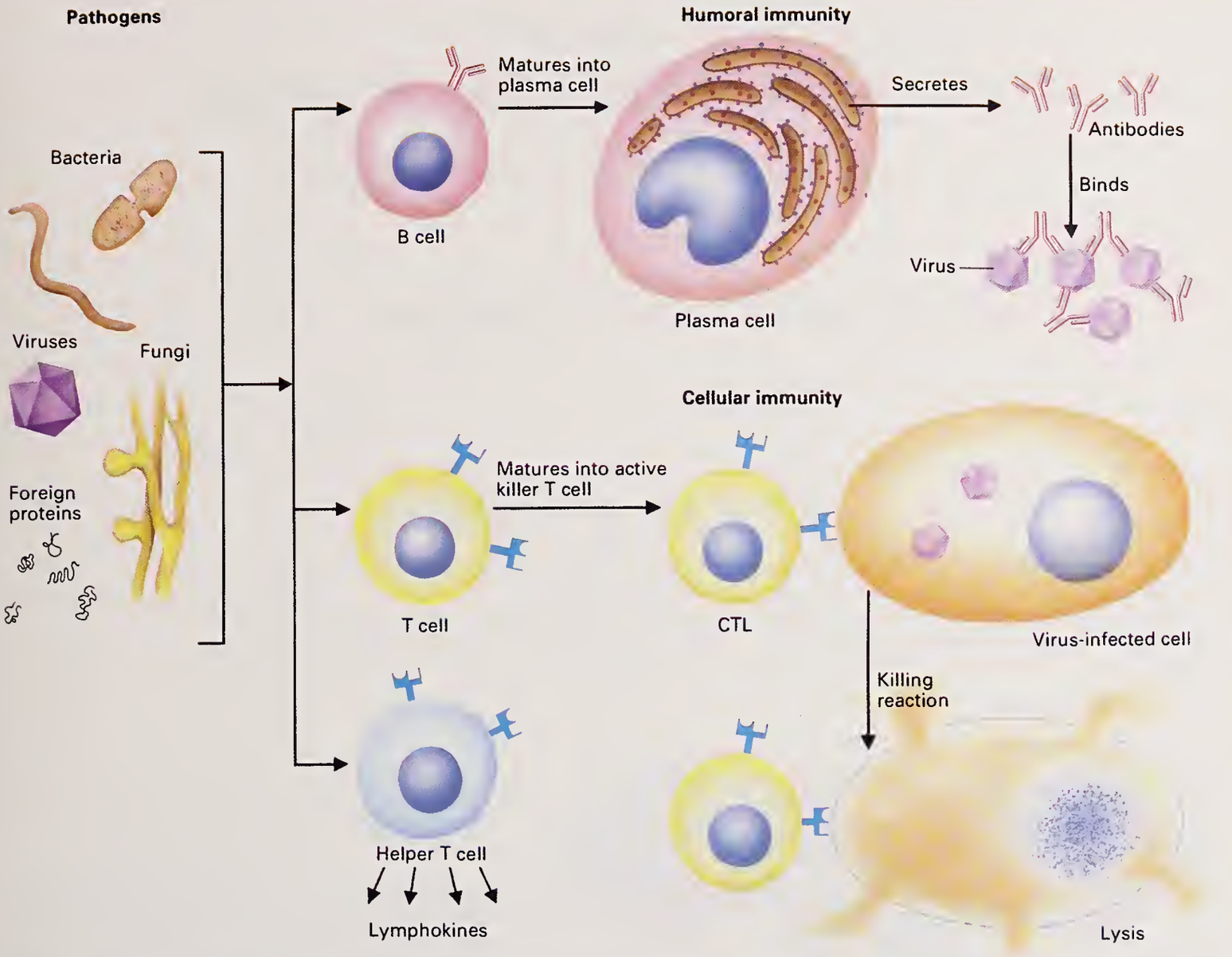


Figure 14. When a pathogen invades the body, the immune system responds with three types of reaction. The cells of the humoral immune system (B cells) secrete antibodies that can bind to the pathogen. Cells of the cellular system (T cells) carry out two major types of functions. One type of T cell (the cytotoxic T cells, CTL) develops the ability to kill pathogen-infected cells. Helper T cells, on the other hand, secrete factors (lymphokines) that stimulate the body's overall response.

From Molecular Cell Biology, *2nd edition. James Darnell, Harvey Lodish, and David Baltimore. p. 1005. Copyright © 1990 by Scientific American Books, Inc. Reprinted with permission by W.H. Freeman and Company.*

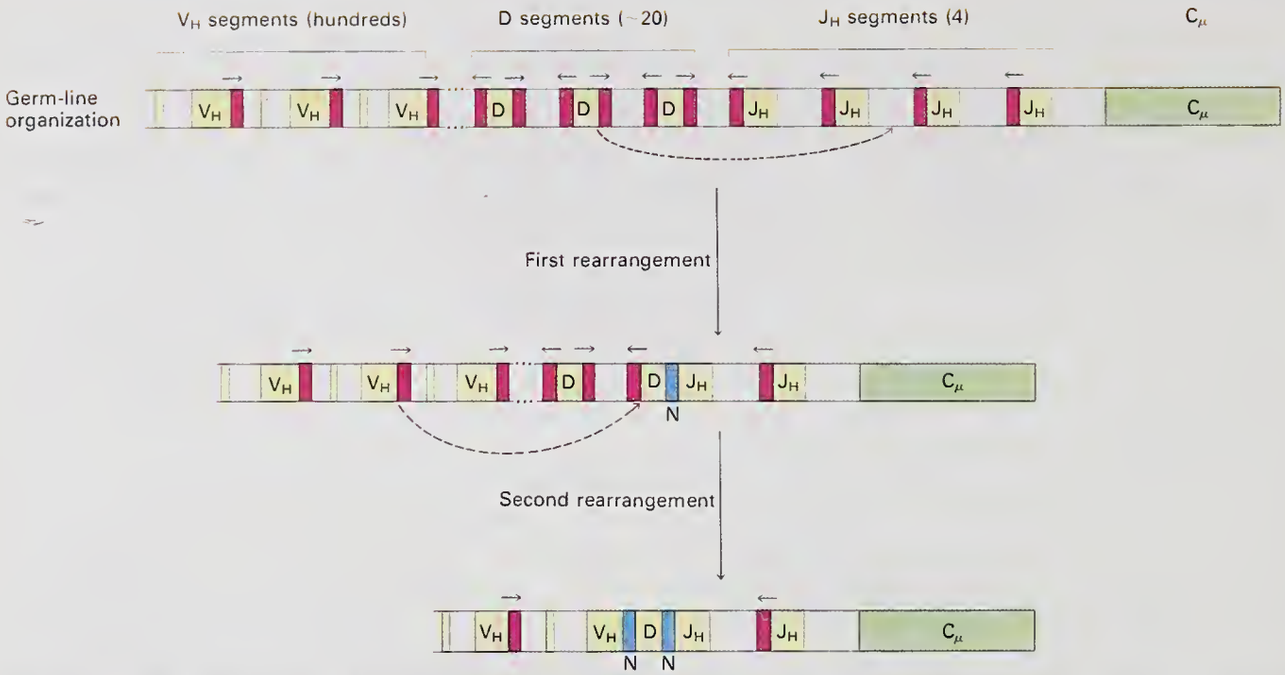


Figure 15. The specificity of antibodies and the receptors on T cells is brought about by a series of complex rearrangements of various gene segments like those shown in this figure.

From Molecular Cell Biology, *2nd edition. James Darnell, Harvey Lodish, and David Baltimore. p. 1027. Copyright © 1990 by Scientific American Books, Inc. Reprinted with permission by W.H. Freeman and Company.*

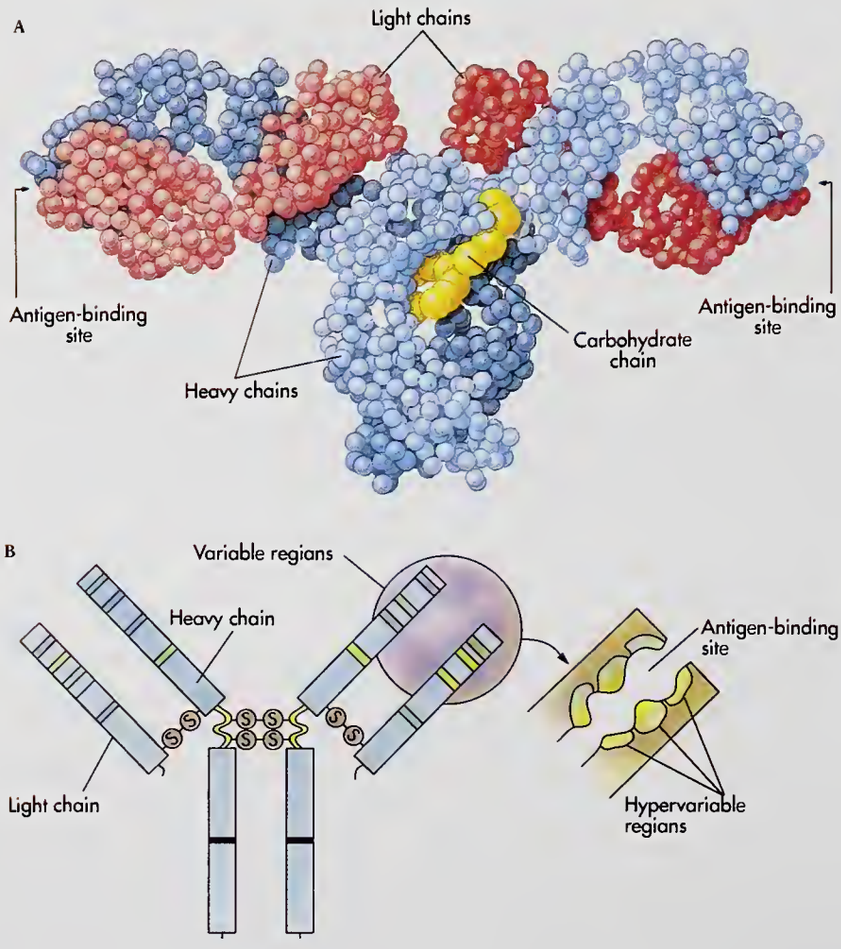


Figure 16. A: A molecular model of an antibody molecule. Each amino acid is represented by a small sphere. The heavy chains are colored blue, the light chains red. The four chains wind about one another to form a Y shape, with two identical antigen-binding sites at the arms of the Y and a tail region that serves to direct the antibody to a particular portion of the immune response.

B: A schematic drawing of an antibody molecule. Each molecule is composed of two identical light (L) chains and two identical heavy (H) chains. Carbohydrate is sometimes associated with the H chain. While the antigen-binding sites are formed by a complex of both H and L chains, the tail region is formed by H chains alone.

From Raven, P.H., and Johnson, G.B. 1988. Understanding Biology. *St. Louis, MO: Times Mirror/Mosby College Publishing, p. 692.*

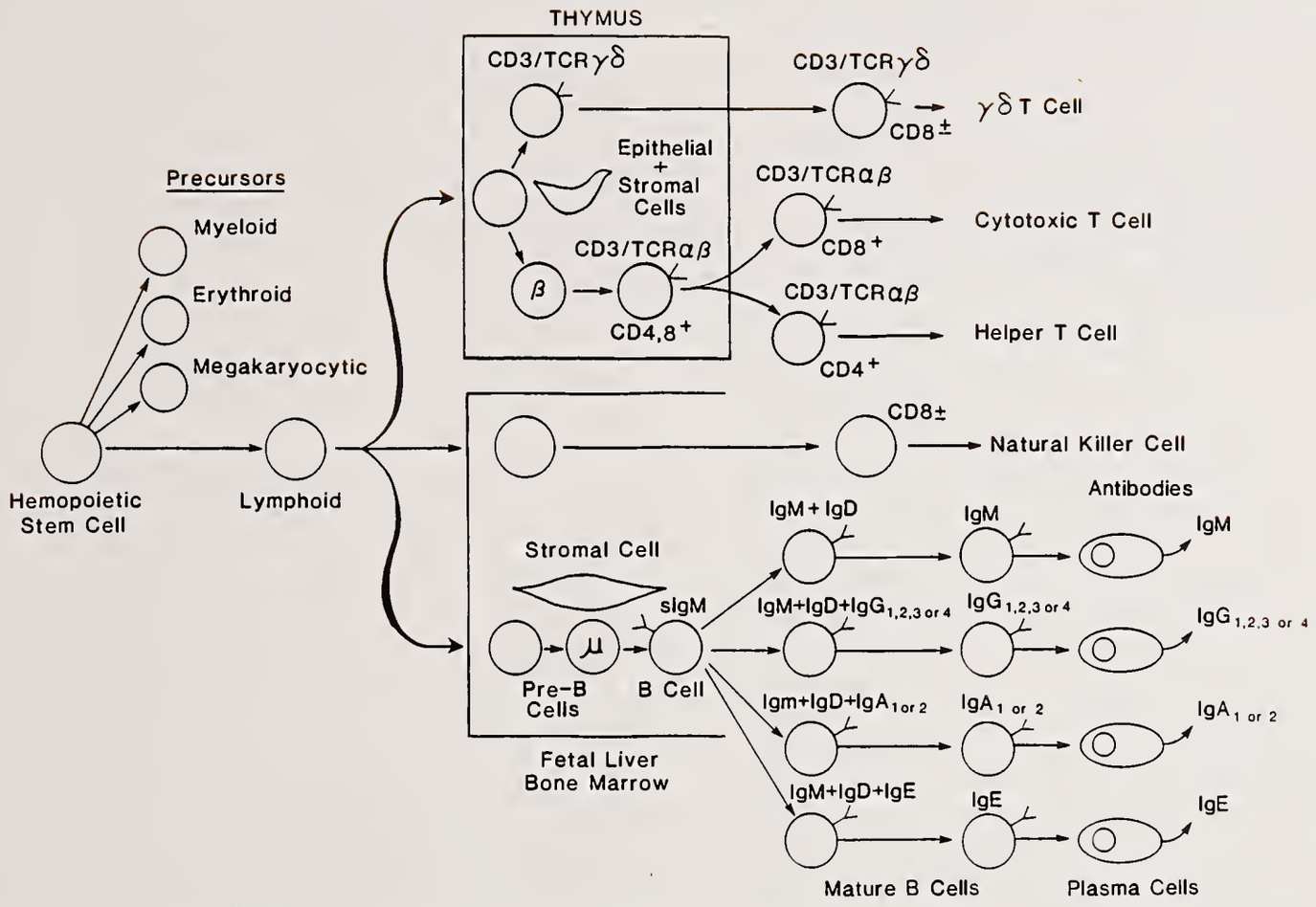
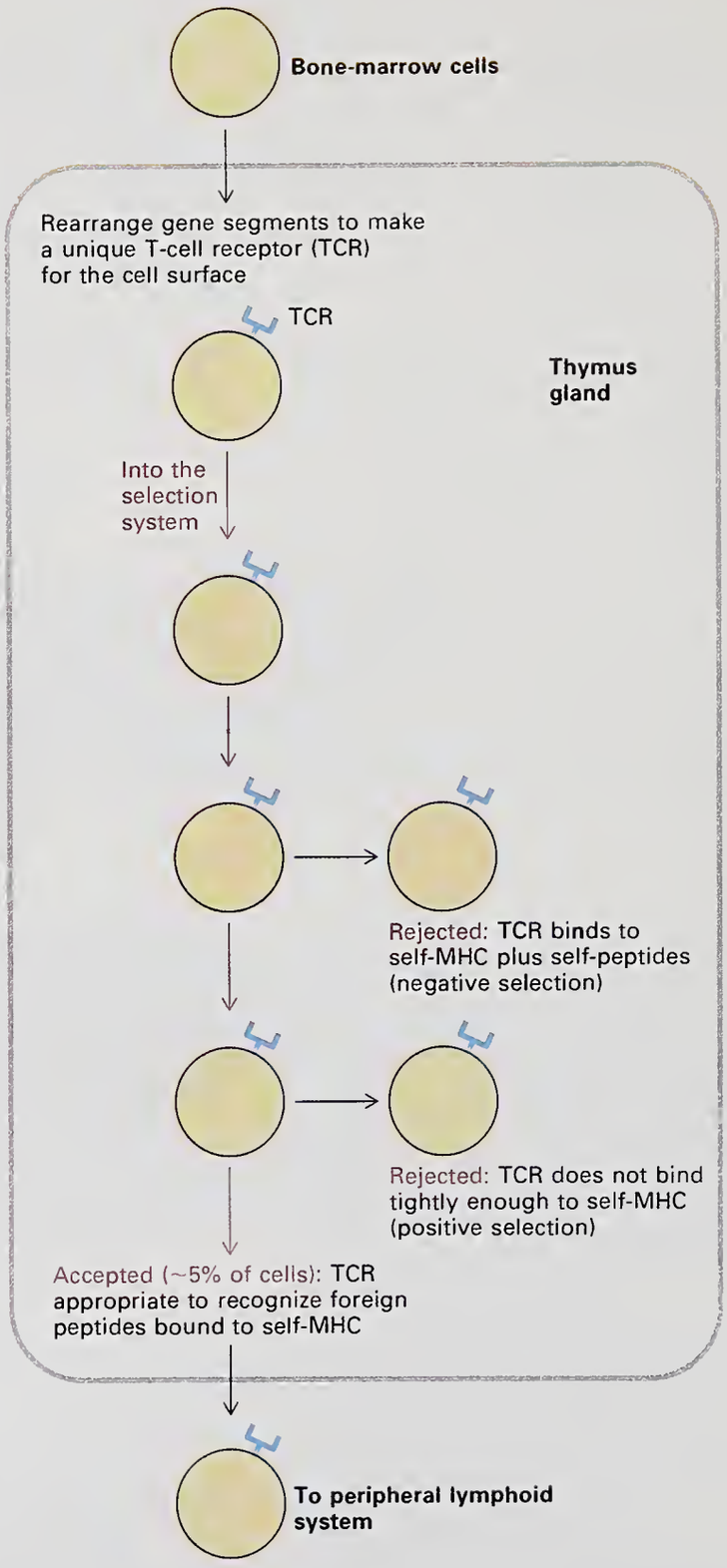


Figure 17. Contemporary model of immune system development in mammals. T and B cells, which form the major recognition and effector limbs of the immune system, are derived from multipotent stem cells that also give rise to the other types of blood cells. Definition of the T cell receptors and accessory cell surface molecules has allowed the identification of two major sublineages of the thymus-derived T cells, each of which expresses a different type of T cell receptor and exhibits specialized function. The B lineage cells, generated in hemopoietic tissues, initially express one class of antibody receptors called immunoglobulin M (IgM), but may then switch to the expression of other antibody classes, each of which has special biological advantages. Note the third lineage of lymphocytes, the natural killer cells. They are the most recently recognized members of the lymphocyte family, but may well prove to be the most phylogenetically ancient elements in this constellation of cells involved in host defense.

Courtesy of Max D. Cooper.

Figure 18. A schematic representation of how the thymus gland selects T cells that are able to recognize foreign but not self proteins.

From Molecular Cell Biology, *2nd edition. James Darnell, Harvey Lodish, and David Baltimore. p. 1037. Copyright © 1990 by Scientific American Books, Inc. Reprinted with permission by W.H. Freeman and Company.*



(Figure 15). Each gene segment exists in several —and in some cases hundreds—of different copies. These segments randomly recombine to form new genes that encode the virtually limitless repertoire of recognition elements. To take just one example, T cells form their receptors by combining a number of different gene sequences: V (**variable**), D (**diversity**), and J (**joining**) segments. From this array any given T cell derives 1 from about 100 possible V segments, 1 from about 6 D segments, and 1 from about 50 J segments to form its so-called α- or **heavy chain**, and about 1 in 20 V, 1 in 2 D, and 1 in 12 J segments to form its **light chain**. The random recombination of V, D, and J segments in the two chains can thus code for literally millions of different possible receptor structures. The receptors on B cells are formed in basically the same way, although the numbers of V, D, and J segments available for selection and recombination differ somewhat. Antibodies (which are secreted by plasma cells) are generated in much the same way as their receptors, and have the same almost unlimited capacity for diversity (Figure 16).

At one time it was thought that antigens were capable of shaping the structure of lymphocyte receptors and antibodies so that the binding sites of receptors and antibodies would mold themselves in some way to fit the shape of the antigen, much as a rubber glove molds itself to fit one's hand. We now know that this "instructional" hypothesis is wrong. Rather, as indicated above, the immune system produces very large numbers of different types of receptors and antibodies, and collectively these can "fit" essentially every possible antigen (Figure 17). Each T or B cell bears only one type of receptor on its surface (although there are thousands of receptor molecules of that given structure on each cell). In the same way, each B cell secretes antibodies of only a single, defined structure. Thus the capacity of the body to respond to an enormous variety of different antigens is due to the existence of an enormous number of different T or B cells, each able to recognize a single antigen (or more commonly, a part of a complex antigen known as an **antigenic determinant**). And when the cell recognizes and binds to an antigen, it responds by proliferating to form a large number of cells of the same type. Such a population of cells, all derived from a single progenitor, is known as a **clone**, and the hypothesis put forward to account for the selective proliferation of lymphocytes of particular receptor type in response to a specific antigen is known as the **clonal selection hypothesis**. This theory, first advanced by Sir F. Macfarlane Burnet, has withstood every test and is rightly viewed as one of the cornerstones of modern immunology.

The intriguing question as to how the cells of the immune system distinguish foreign molecules from those on the surface of the cells of their own host was also addressed by Burnet. He suggested—and there is now a large body of evidence to support this view—that lymphocytes that recognize the body's own tissues (so-called **self antigens**) are selectively eliminated during early development—a process known as **clonal deletion**. The mechanisms responsible for clonal deletion of T and B cells are still under investigation, but at least for the T cells it appears that during development the thymus may actively select, for export to the rest of the immune system, only those T cells that are capable of functioning in the host (Figure 18). These "useful" T cells are allowed to survive and mature, while the potentially harmful cells die and are removed. This process results in the death of about 90 percent or more of the T cells that are initially formed. The bone marrow selection of B cells for survival may be equally stringent. Although this seems an astonishingly wasteful process, comparable cell deaths are known to be a rather common feature in the development of virtually all organs and of all multicellular organisms.

For reasons that remain to be elucidated, in some conditions—commonly referred to as **autoimmune disorders**—the immune system may mistakenly mount an attack on components of the host organism's own cells. For example, the neurological condition myasthenia gravis involves the production of circulating antibodies directed against the receptor molecules on the surfaces of muscle cells that normally enable them to respond to the release of the neurotransmitter acetylcholine from the motor nerves. When the receptor molecules are damaged or destroyed, there is a progressive loss of neuromuscular control, and if the respiratory muscles are involved the condition may be fatal. Similarly, type I or juvenile-onset diabetes is now known to be due to the combined attack of T cells and antibodies directed against the β-cells of the pancreas that normally produce insulin, the hormone that regulates sugar metabolism.

Another topic of considerable current interest in immunology concerns **antigen presentation**. We now know that this involves a complex set of genes called the **major histocompatibil-**

ity complex (MHC). The membrane proteins encoded by these genes (of which there are two types called class I and class II) are able to selectively bind short segments of partially digested protein antigens, termed peptides. These peptides, arising from protein breakdown inside the cell, reach the surface of the cell together with the relevant MHC molecule. Recent x-ray crystallographic studies indicate that the peptide antigen is lodged within a distinctive groove on the outer surface of the MHC molecule, where it can be detected by a lymphocyte bearing the appropriate receptor. The receptors on T cells are specialized to recognize antigens only in the form of such a peptide:MHC complex. There are accessory molecules on the surfaces of T cells, called CD4 and CD8, that are selectively expressed on cells that recognize antigens presented by MHC II and MHC I molecules, respectively. Functionally these accessory molecules form part of the T cell receptor for peptide:MHC complexes by binding to both the MHC molecule and the T cell receptor. For this reason CD4 and CD8 are sometimes called **co-receptors**.

Like the antigen receptors on B and T cells, MHC molecules show considerable diversity. However, this diversity is not due to the recombination of different gene segments but rather to genetic polymorphism. There may be as many as 100 different genetic sequences (**alleles**) at a single MHC locus, and T cells are selected during development only if they can recognize peptides presented by self MHC molecules. How this occurs is unknown, but its role in T cell selection is obviously fundamental.

The CD4-bearing T cells (also called T4 cells, helper lymphocytes, or $CD4^+$ cells) have become widely known because of their role in the development of **AIDS (acquired immune deficiency syndrome)**. The virus that causes AIDS —the human immunodeficiency virus (or HIV)—selectively invades these cells, because the $CD4^+$ molecule fortuitously serves also as a specific receptor for the virus (Figure 19). On entering the $CD4^+$ T cells, the genetic material of the virus (which is formed of RNA) is reverse transcribed into DNA, and this, in turn, becomes integrated into the T cell's own genome. In this way the virus subverts the cell's genetic machinery and, when activated, the cell produces more and more virus, until ultimately the cell is killed. When the cell dies it releases virus into the bodily fluids, where it is free to invade other $CD4^+$ T cells, and the whole process may be repeated until the entire T cell population is effectively depleted. Since, as we have seen, T helper cells are essential for mounting both cell-mediated and humoral immune responses, patients with AIDS become progressively more vulnerable to all forms of infection and commonly succumb to opportunistic infections that would normally be easily overcome.

A second important component of the immune system is the **complement system**, which consists of a complex series of proteins in the serum and in cell membranes (Figure 20). These proteins perform essential roles in the immune response to foreign organisms such as bacteria and viruses, and in the response to tumors. Deficiencies of any of the complement proteins may lead to diseases, including those that involve infection, hemolysis of red blood cells, or autoimmune diseases such as systemic lupus erythematosus.

The complement system is activated by two general mechanisms. First, antibodies (Ab) can activate complement when they bind their antigen (Ag). In addition to this so-called classical pathway there is an alternative pathway that is continuously active at a low level marking foreign organisms for which there are no preformed antibodies available.

In addition to these roles, complement proteins help to regulate the immune system by another mechanism. This involves the interaction of specific activated complement protein fragments with receptors, or binding proteins, that are on the surface of immune system cells. These receptors allow for communication with the interior of the cell, and their activation leads to a change in the function or fate of the cell.

Overall, the complement system plays a fundamental role in normal or abnormal immune responses. Current study in this area is directed toward understanding not only the molecular mechanisms of complement activation and regulation but also the general effects on the immune response of experimentally altering complement function.

The devastating consequences of AIDS, congenital immunodeficiency disorders, and the frequent rejection of transplanted organs have made the public increasingly aware of the importance of the immune system in medical practice. The development of immunosuppressive drugs has gone a long way toward overcoming the problem of tissue rejection, and there is now considerable interest in the possible develop-

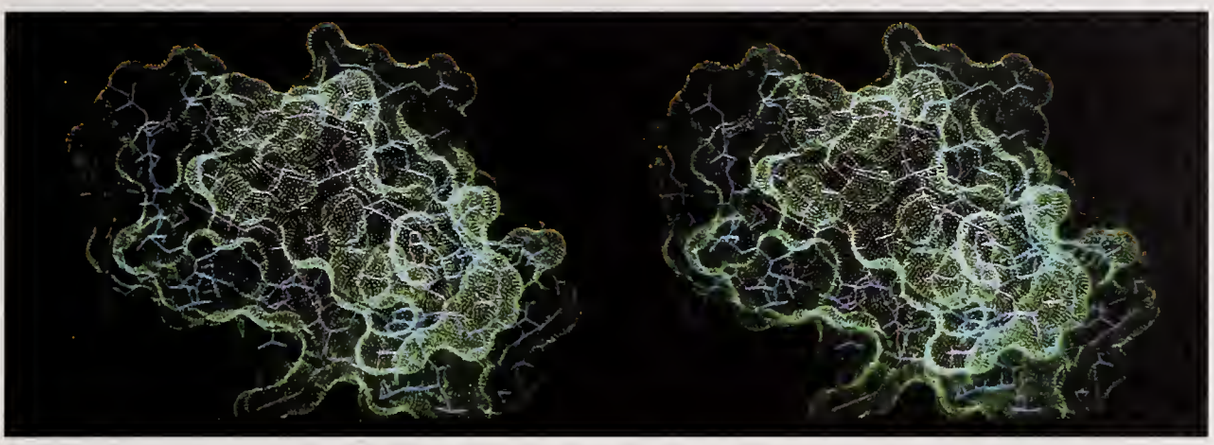


Figure 19. Stereoview of a region on the macromolecule CD4, the T cell receptor for HIV. CD4 contains four tandem immunoglobulin-like domains, D1–D4. The view represents D1's major binding region for the virus coat glycoprotein gp120. Atoms of residues 41–59 are shown in red. They are enveloped by the surface in contact with a water molecule probe. This figure was produced with the display program QUANTA.

Reprinted by permission from Ryu, S.-E., Kwong, P.D., Truneh, A., Porter, T.G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N.-h., Axel, R., Sweet, R.W., and Hendrickson, W.A. Nature *348:419–426. Copyright © 1990 Macmillan Magazines Ltd.*

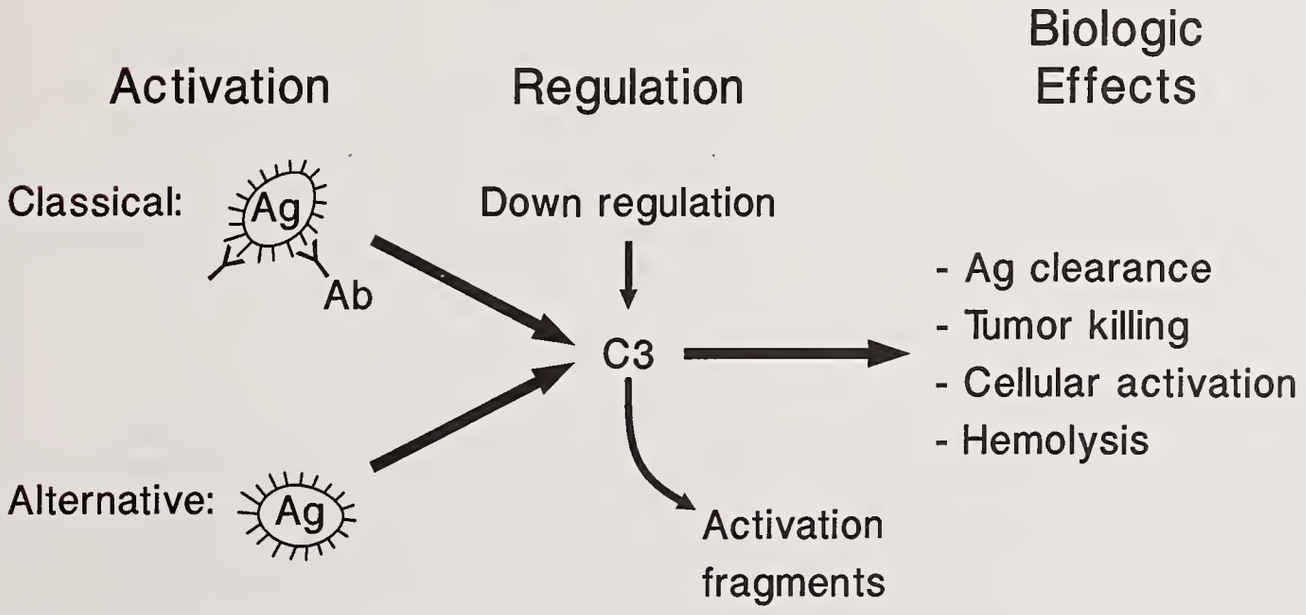


Figure 20. The complement system. Ag, antigen; Ab, antibodies. Courtesy of V. Michael Holers.

ment of vaccines to limit the spread of HIV. Many of us remember how some 30 years ago poliomyelitis was to all intents and purposes eliminated in this country by the development of the Salk and Sabin vaccines, and we are all conscious of the fact that many illnesses such as measles, rubella, whooping cough, and even smallpox have been brought under control. But it is still not widely appreciated that the immune system is itself subject to a number of serious disorders such as lymphoma and leukemia. Our ability to deal with these malignant conditions is still very limited, but we are beginning to understand what may cause them. While these disorders present the most urgent challenges to clinical immunology, even relatively minor allergic disorders continue to pose problems both for the practicing physician and for the patients who suffer from them. Many of the reports in this volume indicate how these and other problems associated with the immune system are currently being addressed.

Investigators in the Immunology Program

Alt, Frederick W., Ph.D.
Atkinson, John P., M.D.
Bevan, Michael J., Ph.D.
Bjorkman, Pamela J., Ph.D.
Bloom, Barry R., Ph.D.
Bottomly, H. Kim, Ph.D.
Chaplin, David D., M.D., Ph.D.
Cooper, Max D., M.D.
Cresswell, Peter, Ph.D.
Davis, Mark M., Ph.D.
Fischer Lindahl, Kirsten, Ph.D.
Flavell, Richard A., Ph.D.
Ghosh, Sankar, Ph.D.
Goodnow, Christopher C., B.V.Sc., Ph.D.
Holers, V. Michael, M.D.
Jacobs, William R., Jr., Ph.D.
Janeway, Charles A., Jr., M.D.
Kappler, John W., Ph.D.
Korsmeyer, Stanley J., M.D.
Leiden, Jeffrey M., M.D., Ph.D.
Littman, Dan R., M.D., Ph.D.
Lob, Dennis Y.-D., M.D.
Marrack, Philippa, Ph.D.
Nussenzweig, Michel C., M.D., Ph.D.
Payan, Donald G., M.D.
Perlmutter, Roger M., M.D., Ph.D.
Peterlin, B. Matija, M.D.
Schatz, David G., Ph.D.
Smale, Stephen T., Ph.D.
Thomas, Matthew L., Ph.D.
Thompson, Craig B., M.D.
Tonegawa, Susumu, Ph.D.
Weiss, Arthur, M.D., Ph.D.
Weissman, Irving L., M.D.
Witte, Owen N., M.D.

Neuroscience Program

Among the most challenging problems in biomedical research are those posed by the human brain. How do we perceive the world around us? How do we learn from past experiences? How do we store and recall information derived from those experiences? How do we determine when to act and what actions to carry out? What is thought? And what are the neural mechanisms that underlie language? In a word, how are all those aspects of our lives that most specifically define our humanity instantiated in the functioning of our brains? The answers to these questions still lie far in the future, but in the past two decades considerable progress has been made in our understanding of some of the cellular and molecular mechanisms involved in brain function. Recognizing this, in 1983 the Institute initiated its Neuroscience Program. Until recently, the Neuroscience Program has been largely focused on the ways in which **nerve cells** conduct signals and communicate with each other and with the effector tissues of the body (such as muscle and gland cells) and on the cellular mechanisms involved in the development of the nervous system.

Modern neuroscience is founded on two fundamental concepts that derive from the late nineteenth and the early years of the twentieth century. The first of these, commonly referred to as the **neuron doctrine**, is that the fundamental functioning units of the nervous system are nerve cells, or **neurons**. Among the cells of the body, neurons are distinguished anatomically by the fact that they all extend processes (some of considerable length) that are of two general types: shorter tapering processes (**dendrites**) that mainly serve to receive information from other cells, and longer processes (**axons**), of more uniform diameter, that serve to transmit information to other parts of the nervous system or to the body at large. The second basic concept is that infor-

mation in the nervous system is principally encoded in a series of signals called **nerve impulses**, or **action potentials**. These are brief, usually all-or-nothing electrical changes in the nerve cell membrane that are propagated along the axons at rates between about 3 and 400 feet per second. A necessary corollary of this concept is the notion that nerve cells communicate this encoded information to each other at specific sites called **synapses**, where the axon of one cell functionally interacts with the dendrites or the bodies of other neurons.

The essential morphological features of neurons were established in the 1870s and 1880s, with the aid of a number of selective staining procedures, notably the metallic impregnation technique developed by the Italian microscopist Camillo Golgi. And the fundamental principle of the neuron doctrine, namely, that nerve cells are anatomically and functionally discrete entities, was convincingly demonstrated around the turn of the century by the great Spanish neurohistologist, Santiago Ramón y Cajal. The biophysical mechanisms responsible for the nerve impulse and for synaptic transmission were established in the early 1950s, principally through the work of Hodgkin, Huxley, Katz, and Eccles (Figures 21, 22, and 23). In brief, activation of a nerve cell results in the successive opening of pores or ion channels along the length of the axon that result in the temporary reversal of the voltage between the inside and the outside of the axon (this transient change in potential is the action potential; see Figure 7). When the action potential reaches the ends of the axon it causes the release of a **neurotransmitter** that diffuses across the microscopic gap between the axon terminal and the postsynaptic cell. The binding of the neurotransmitter to specialized receptors in the membrane of the postsynaptic cell in turn triggers a response in that cell which may either be the opening of an ion channel or the activation of a second intracellular messenger in the cell. In either case the binding of the transmitter to the receptor is reflected in the generation of a graded voltage change across the membrane of the postsynaptic cell, called a **synaptic potential**. Depending on the nature of the transmitter receptor, the response may be either excitatory or inhibitory; i.e., the postsynaptic cell may either be activated or rendered less likely to discharge an impulse. Finally, the released neurotransmitter is either broken down by a specific enzyme within the synaptic cleft, or taken up by selective transport mechanisms into the axon terminal (where it can be reutilized) or into the surrounding nonneural (**glial**) cells.

In the past 10 years we have learned a good deal about the molecular mechanisms involved in both impulse conduction and synaptic transmission, largely as the result of the successful cloning of the genes for a number of the ion channels involved (e.g., for Na^+, K^+, and Ca^{2+}) and for many neurotransmitter receptors like those for acetylcholine, glutamate, γ-aminobutyric acid (GABA), serotonin, norepinephrine, dopamine, and various neuropeptides. From the nucleotide sequence of these genes it has been possible not only to deduce the primary amino acid sequence of the channel or transmitter proteins (and from this to infer the probable arrangement of the relevant protein in the membrane) but also to generate hybridization probes to identify other related channels or receptors. And using some of the established techniques of genetic engineering, like site-directed mutagenesis, it has been possible in some cases to establish the regions within the channel molecules that are sensitive to changes in voltage (Figure 24), or the ligand-binding and second messenger–activating domains of receptors. One example will suffice to demonstrate the importance of this approach to our understanding of these fundamental processes.

It has been known for almost 50 years that the relatively simple molecule **acetylcholine** is the transmitter at the junctional region between motor nerve fibers and muscle cells and also at certain synapses in the brain and spinal cord. With the discovery in the 1970s that the clinical condition **myasthenia gravis** (previously discussed in the immunology section) is caused by circulating antibodies directed against the receptor for acetylcholine in the muscle membrane, a major effort was mounted to purify and biochemically characterize the acetylcholine receptor (AChR). This work served to establish that the AChR consists of five subunits: two designated α, and one each called β, γ, and δ.

In the early 1980s Heinemann, Patrick, and their colleagues succeeded in cloning the genes for the α-subunit, and in 1983 Numa and his coworkers presented the complete nucleotide sequences encoding all four kinds of subunits. From these sequences we gained several important insights. First, the four subunits showed a high degree of homology, which suggested that their genes were probably derived—by duplication and divergence—from a single ancestral

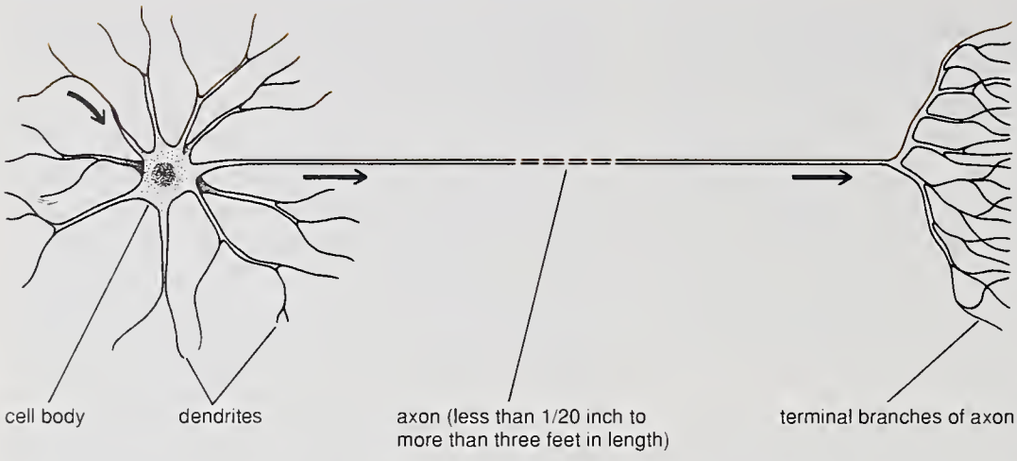


Figure 21. A schematic diagram of a typical neuron. The arrows indicate the direction in which nerve signals are conveyed. The largest axons in the human brain and spinal cord extend for about three feet and have a diameter of less than 1/1,000th of an inch. Many axons are covered by an insulating layer known as the myelin sheath. The myelin sheath is interrupted at intervals known as nodes of Ranvier.

From Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. 1989. Molecular Biology of the Cell, *2nd edition. New York: Garland, p. 1061.*

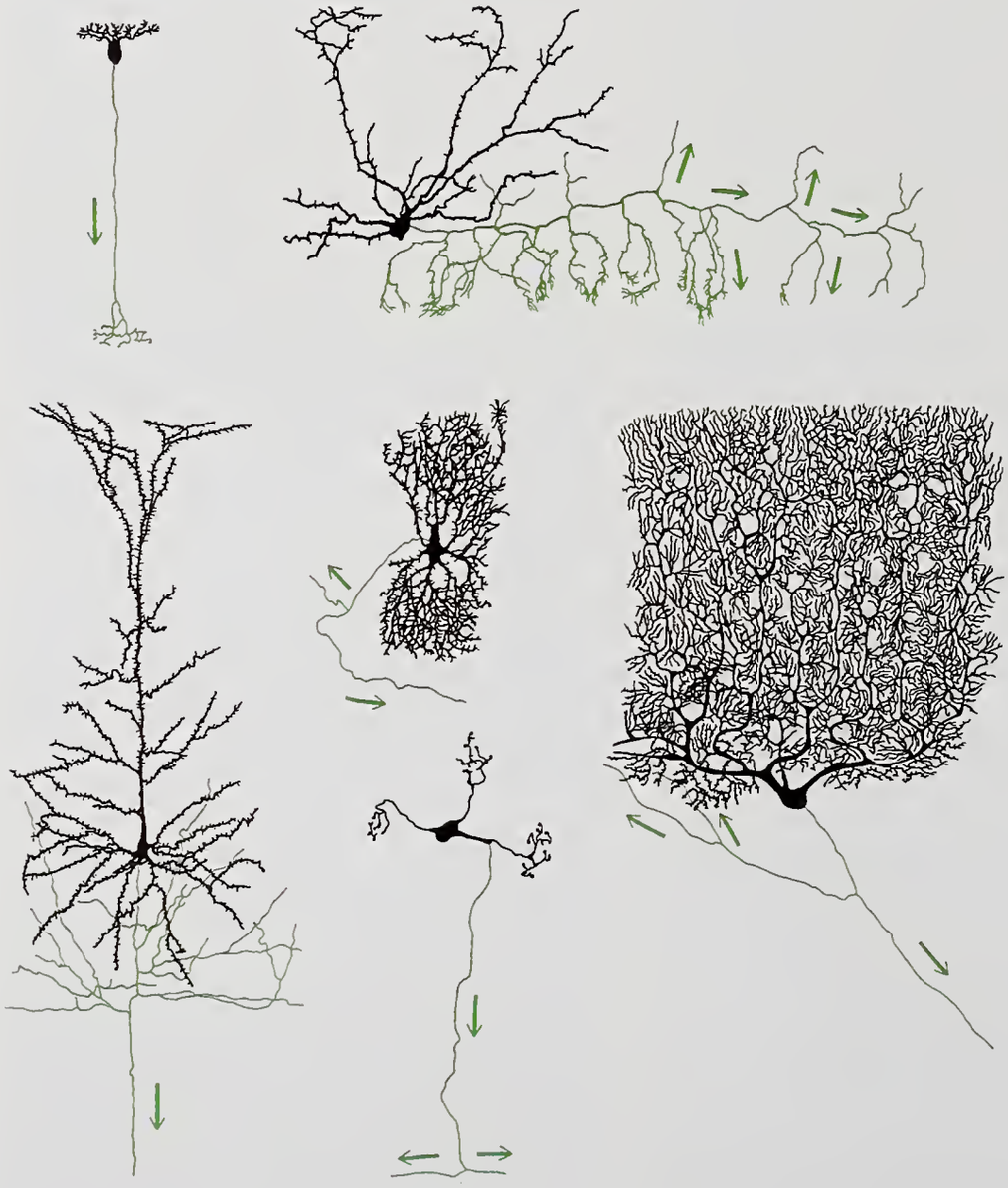
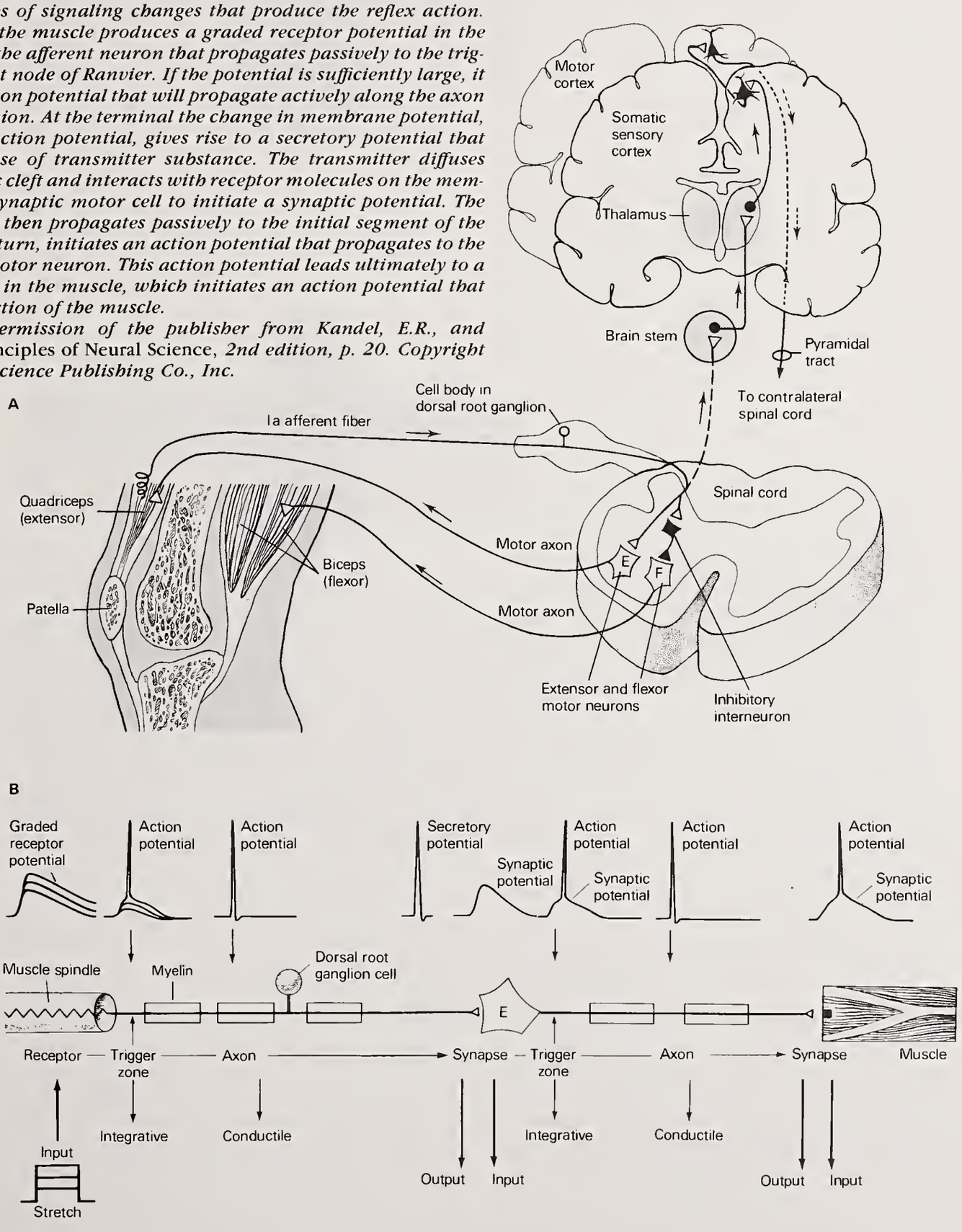


Figure 22. A few of the many types of neurons in the vertebrate nervous system. (S. Ramón y Cajal, Histologie du Système Nerveux de l'Homme et des Vertébrés. *Paris: Maloine, 1901–1911; reprinted, Madrid: C.S.I.C., 1972.)*

From Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. 1989. Molecular Biology of the Cell, *2nd edition. New York: Garland, p. 1061.*

Figure 23. A: The anatomical arrangement of a typical reflex, in this case the knee jerk. Each cell actually represents a population of many neurons. Information about stretch of the quadriceps femoris muscle is conveyed by afferent neurons to several loci within the central nervous system. In the spinal cord, afferent neurons act directly on the motor neurons to the quadriceps and, by means of inhibitory interneurons, indirectly on the motor neurons to the antagonistic muscle, the biceps. Both of these actions combine to produce the coordinated expression of the reflex behavior. In addition, information is conveyed to higher regions of brain to update them about the information coming into the nervous system and about the behavior that is being generated. These higher centers, in turn, can act to modify the reflex behavior.

B: The sequences of signaling changes that produce the reflex action. Graded stretch of the muscle produces a graded receptor potential in the muscle spindle of the afferent neuron that propagates passively to the trigger zone at the first node of Ranvier. If the potential is sufficiently large, it will trigger an action potential that will propagate actively along the axon to the terminal region. At the terminal the change in membrane potential, produced by the action potential, gives rise to a secretory potential that leads to the release of transmitter substance. The transmitter diffuses across the synaptic cleft and interacts with receptor molecules on the membrane of the postsynaptic motor cell to initiate a synaptic potential. The synaptic potential then propagates passively to the initial segment of the axon, where it, in turn, initiates an action potential that propagates to the terminals of the motor neuron. This action potential leads ultimately to a synaptic potential in the muscle, which initiates an action potential that causes the contraction of the muscle.



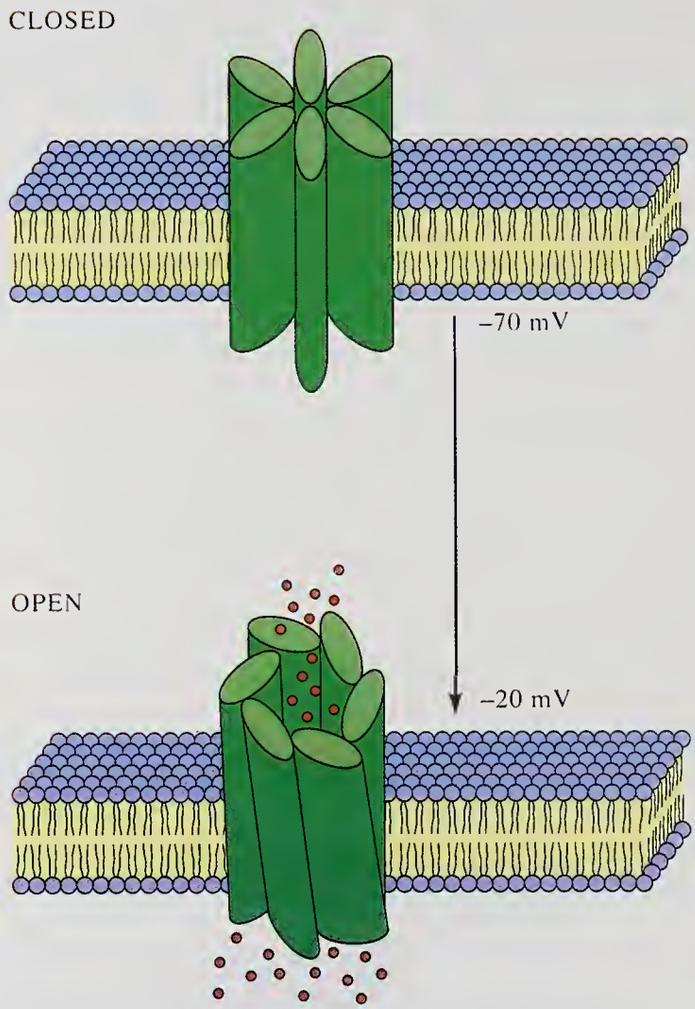


Figure 24. A local reduction of the voltage differential may induce a sodium or potassium channel to change conformation from one allosteric form to another, thus opening the channel and leading to the free passage of ions. This change is presumably caused by key electrically charged amino acids in the channel protein, which shift their orientation in response to the changed electrical field.

From Discovering Enzymes, *by David Dressler and Huntington Potter, p. 233. Copyright © 1991 by David Dressler. Reprinted with permission by W.H. Freeman and Company.*

gene. Second, the similarities in the predicted amino acid sequences suggested that the subunits are probably arranged to form a central pore or channel between them. Third, the presence in each subunit of four hydrophobic regions, each about 20 amino acids long, immediately suggested the probable disposition of the subunits, with four transmembrane domains (M1, M2, M3, and M4) and intervening intra- and extracellular linking segments. More recent work has identified a fifth subunit type ϵ and has established the precise location of the acetylcholine-binding site and the existence of a family of nicotinic AChRs in the central nervous system; much has also been learned about the regulation of the receptor during muscle development and after denervation, and about the process of receptor desensitization.

The nicotinic AChR belongs to a large class of neurotransmitter receptors that operate by selectively opening ion channels. Another, somewhat larger class of receptors acts through second messengers. For example, the adrenergic receptors, which are responsible for controlling a number of vital functions such as heart rate and blood pressure, act through the intermediary of a class of so-called **G proteins** to activate the enzyme adenylate cyclase and increase the intracellular level of the important second messenger, cAMP. The second messenger, in turn, usually acts by stimulating protein kinases that modify (by adding phosphate groups) other proteins, including ion channels and proteins that regulate gene expression in the responding cell.

The regulation of gene expression by synaptically mediated second messenger systems has become one of the most active areas for research in molecular neuroscience. Whereas neurotransmitters usually result in changes that have a time course measured in the millisecond to second range, many of the most intriguing phenomena in neuroscience are those that occur over periods of hours, days, and even months or years. Recent work has demonstrated that in addition to their more or less immediate and short-lasting effects, under appropriate conditions (such as those we commonly associate with learning and memory) neurotransmitters may, through second messengers, activate a number of transcriptional regulatory proteins that "turn on" various classes of genes. These, in turn, may regulate the expression of yet other genes and thus unleash a complex cascade of events within the responding nerve cell, modifying its growth and altering its responsiveness to later neurotransmitter activation over long periods of time.

One of the major beneficiaries of the application of the new genetics to the nervous system has been the field of developmental neuroscience. Indeed it is no exaggeration to say that since the late 1970s this field has been transformed from an essentially descriptive science into one in which, for the first time, mechanistic explanations are emerging to account for the growth of nerve cells and their processes, for the deployment of cells into peripheral ganglia and within cortical layers or nuclear groups in the central nervous system, for the formation of specific patterns of connections, and for the elimination of redundant cells and inappropriate connections.

Because of the complexity and inaccessibility of the mammalian central nervous system, until recently much of the most definitive work on neural development has been carried out in simpler forms such as the nematode *C. elegans* and the fruit fly *Drosophila*. It is difficult to summarize the broad sweep of this work, except to say that it has served to clarify the genetic mechanisms that determine the distinct front-to-back and top-to-bottom organization of all developing organisms, that determine not only which cells will become neurons but also how many neurons will be generated and what type they will be (e.g., sensory cells, interneurons, or motor cells), and that determine finally whether the neurons that are initially formed will survive. In some instances it is clear that the character or **phenotype** of the nerve cells is determined by their lineage; in other cases cell-cell interactions are more important, and the nature of the signals that developing cells transmit to their neighbors is currently being elucidated. Of special importance are the molecules on the surfaces of cells that enable them to recognize and aggregate with other cells of like kind or enable them to migrate along other cells or across territories filled with extracellular matrix materials. While much remains to be discovered, the first fruits of this harvest hold great promise for future progress in this important field.

Finally, no account of progress in molecular neuroscience would be complete without reference to the striking developments in our understanding of the basis of some of the major genetic disorders that affect the nervous and related muscular systems. Perhaps the most striking of these developments has been the cloning of the gene

for Duchenne and Becker muscular dystrophy. These are X-linked recessive disorders that, in the more severe (**Duchenne**) form lead inexorably from muscular weakness to muscular atrophy and finally death. The extreme size of this gene (it comprises about 1 percent of the X chromosome and almost 0.1 percent of the total human genome) renders it especially vulnerable to mutation, and in many of the identified mutations, the protein encoded by the gene, **dystrophin** (which appears to be critical for coupling muscle excitation and contraction), is either absent or markedly deficient.

Some years ago the general location of the gene responsible for the severe neurological disorder known as Huntington's disease (HD) was determined using RFLPs, as discussed in the section on genetics. Although the HD gene itself has so far eluded us, there is every reason to be optimistic that within a year or two it will be identified and cloned. In the meantime, the relevant RFLP has provided a useful marker for identifying carriers of the disordered gene. There is similarly reason for optimism that in the near future the genetic basis for the two major affective disorders, manic depression and schizophrenia, will be elucidated. Careful studies of family histories and of identical twins raised apart have clearly established that both illnesses have an important genetic component, and while neither is probably due to a single genetic mutation, RFLP analysis and other genetic approaches should reveal the genes involved. That such complex behavioral disorders might yield to this type of approach was unthinkable only a decade ago; as noted in the section on genetics, nothing serves to emphasize more dramatically the power of the new genetics or the exciting possibilities it portends.

The dramatic advances in cellular and molecular neuroscience should not obscure the fact that the distinctive role of the nervous system in the economy of an organism is its capacity to integrate sensory information that is received (both from within and from outside the body) and to organize it into patterns of behavior that allow the organism to respond to changes in its environment in appropriate ways. And for human beings, it is through the nervous system that we learn both from personal experience and the accumulated wisdom of previous generations not only how to survive but how to enjoy and profit from the richness of mental experience and meaningful social interactions. To understand all this we will need to learn a great deal more about the activities of large populations of nerve cells, about the computational capacity of complex neural networks, and about the extraordinary ways in which the human brain, with its 100 billion or more neurons and its more than 1 trillion synapses, receives and processes information. We are at the threshold of being able to understand the logic of the simplest organisms at both the molecular and systems levels; the exploration of human behavior in these same terms stands as perhaps the greatest challenge to modern science.

Investigators in the Neuroscience Program

Adams, Paul R., Ph.D.
Aldrich, Richard W., Ph.D.
Anderson, David J., Ph.D.
Artavanis-Tsakonas, Spyridon, Ph.D.
Axel, Richard, M.D.
Corey, David P., Ph.D.
De Camilli, Pietro, M.D.
Evans, Ronald M., Ph.D.
Goodman, Corey S., Ph.D.
Horvitz, H. Robert, Ph.D.
Huganir, Richard L., Ph.D.
Hurley, James B., Ph.D.
Jahn, Reinhard, Ph.D.
Jan, Lily Y., Ph.D.
Jan, Yuh Nung, Ph.D.
Jessell, Thomas M., Ph.D.
Kandel, Eric R., M.D.
Lerner, Michael R., M.D., Ph.D.
Miller, Christopher, Ph.D.
Movshon, J. Anthony, Ph.D.
Reed, Randall R., Ph.D.
Reichardt, Louis F., Ph.D.
Rosenfeld, Michael G., M.D.
Rubin, Gerald M., Ph.D.
Sakmar, Thomas P., M.D.
Scheller, Richard H., Ph.D.
Sejnowski, Terrence J., Ph.D.
Siegelbaum, Steven A., Ph.D.
Steller, Hermann, Ph.D.
Stevens, Charles F., M.D., Ph.D.
Struhl, Gary, Ph.D.
Südhof, Thomas C., M.D.
Tanabe, Tsutomu, Ph.D.
Tsien, Roger Y., Ph.D.
Yau, King-Wai, Ph.D.
Yellen, Gary, Ph.D.
Ziff, Edward B., Ph.D.
Zipursky, S. Lawrence, Ph.D.
Zuker, Charles S., Ph.D.

Structural Biology Program

The primary goal of structural biology is to understand, in atomic detail, the three-dimensional architecture of proteins, protein assemblies, and the complexes formed by proteins that interact with RNA and DNA. Underlying this approach is the belief that fundamental insights into the functional role of biologically interesting molecules can best come from understanding the forms of the molecules themselves. As Francis Crick, the co-discoverer of the double-helical structure of DNA (see Figure 9), remarked: "to understand function it is essential to study structure." It was with this in mind that in 1985 the Institute made a substantial commitment to develop a new Program in Structural Biology.

At present, **x-ray crystallography** is the most powerful approach for visualizing the three-dimensional structures of large molecules (commonly called **macromolecules**). An essential prerequisite for x-ray analysis is the availability of crystals of the molecule or molecular complex that are suitable for recording the diffraction of x-rays. The production of crystals, in turn, requires chemically homogeneous preparations. Moreover, molecules that are not spatially uniform (too "floppy") must be broken down or molecularly dissected into defined and rigid components. For example, to crystallize antibody molecules, it was important first to cleave them into their principal fragments, Fab and Fc, because these pieces are normally connected by a flexible hinge. And because of antibody diversity (described in the section on immunology), it became necessary to study Fab fragments from monoclonal immunoglobulins. A continuing challenge to structural biologists is the development of strategies for crystallizing membrane proteins—by solubilizing them with detergents, dissecting them into pieces, or altering them by mutation.

Genetic engineering has transformed structural biology. This approach, which makes it possible to produce large quantities of pure proteins, also allows an investigator either to choose a suitable fragment for study or to modify genetically the molecule to be crystallized. Other methodological advances in crystallography itself are transforming the field by extending the range of problems that can be tackled routinely.

There are essentially four stages in determining a structure by x-ray diffraction analysis (Figures 25 and 26): 1) diffraction experiments (data collection); 2) complex computations that produce, in effect, an image of the molecule(s) in the crystal; 3) interpretation of the computed image in terms of a molecular model; and 4) refinement of the model by further computation. **Synchrotron x-ray sources**, which are a thousand or more times stronger than conventional laboratory x-ray generators, are making it possible to study structures that could not previously be solved. (HHMI is currently developing a synchrotron resource for use by the biological community at the National Synchrotron Light Source at Brookhaven National Laboratory on Long Island.) Recent examples from HHMI laboratories are the human class I major histocompatibility antigen and the DNA virus SV40. At the same time, position-sensitive x-ray detectors have greatly extended the applications of conventional radiation sources. Lastly, novel computational methods have made the production of a molecular image (**phase determination**) less dependent on extensive ancillary data from heavy-atom modified crystals and have made refinement of models less cumbersome and more objective.

In the 1950s and 1960s, x-ray crystallography revealed the structures of the first biologically important molecules, including DNA, hemoglobin, and insulin. In the 1970s it revolutionized the field of enzymology by making it possible to visualize directly the active sites of enzymes. In the 1980s it made comparably far-reaching contributions to virology, immunology, and membrane biology by revealing the structures of viruses, antibodies, and a photosynthetic reaction center. What can we expect in the 1990s? It seems reasonable to predict the following:

1. *Structures of different classes of proteins or protein/nucleic-acid complexes.* Three decades of biological crystallography have left several major areas unexplored. For example, we have yet to know what any of the major proteins of the cytoskeleton and of cellular motility look like (actin, myosin, tubulin, and so forth). We have yet to visualize any of the membrane receptors referred to above, and we have yet to see an ion channel, a ribosome, an RNA polymerase, or a ribozyme; and, with the exception of transfer RNAs (tRNAs), little is known of the three-dimensional structure of most RNAs and RNA-protein complexes. Progress toward some of these goals is reported in this volume; others will no doubt be achieved before long, as more and more workers are drawn into the field and as new techniques are developed.

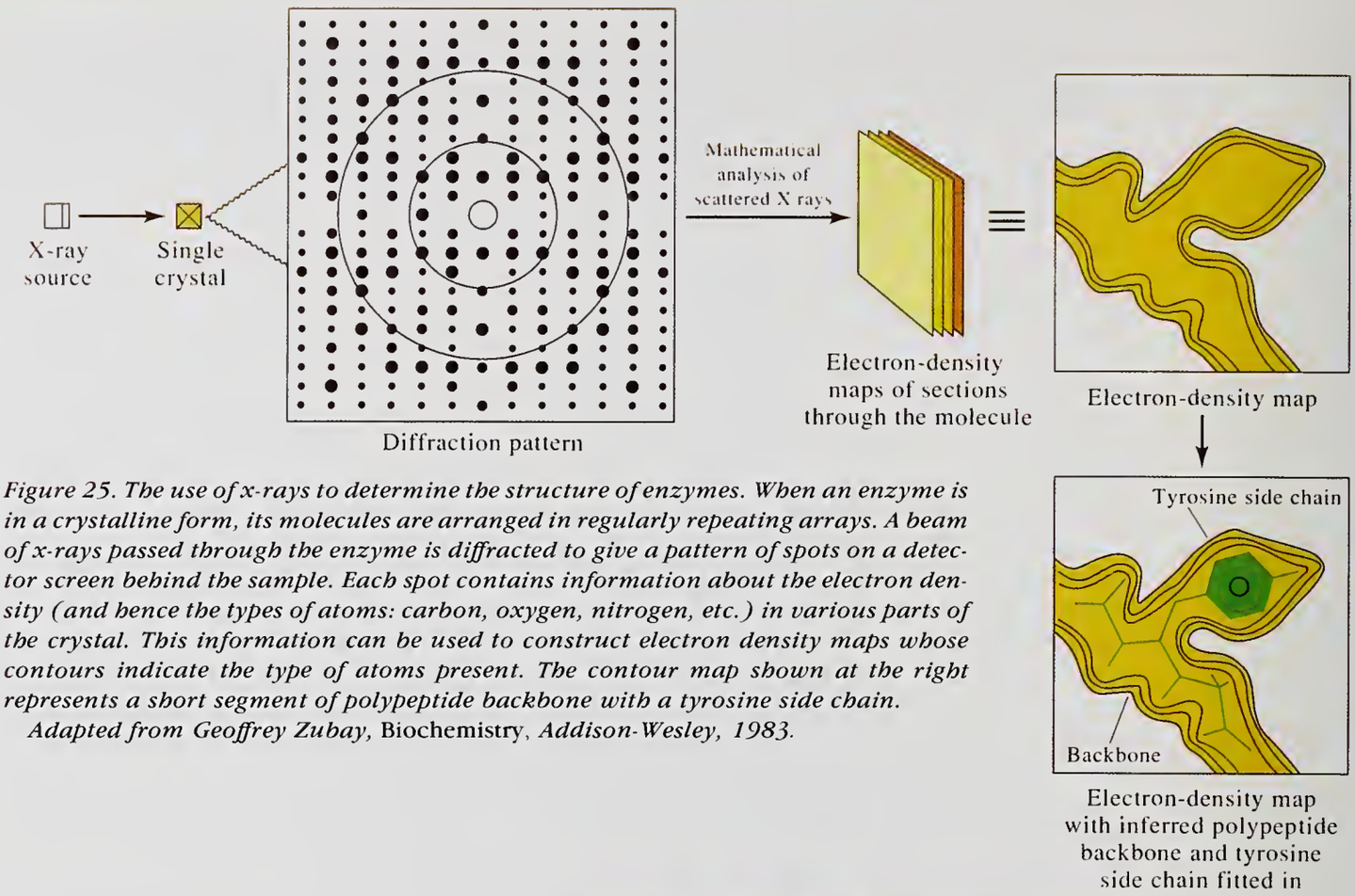


Figure 25. The use of x-rays to determine the structure of enzymes. When an enzyme is in a crystalline form, its molecules are arranged in regularly repeating arrays. A beam of x-rays passed through the enzyme is diffracted to give a pattern of spots on a detector screen behind the sample. Each spot contains information about the electron density (and hence the types of atoms: carbon, oxygen, nitrogen, etc.) in various parts of the crystal. This information can be used to construct electron density maps whose contours indicate the type of atoms present. The contour map shown at the right represents a short segment of polypeptide backbone with a tyrosine side chain.

Adapted from Geoffrey Zubay, Biochemistry, *Addison-Wesley, 1983.*

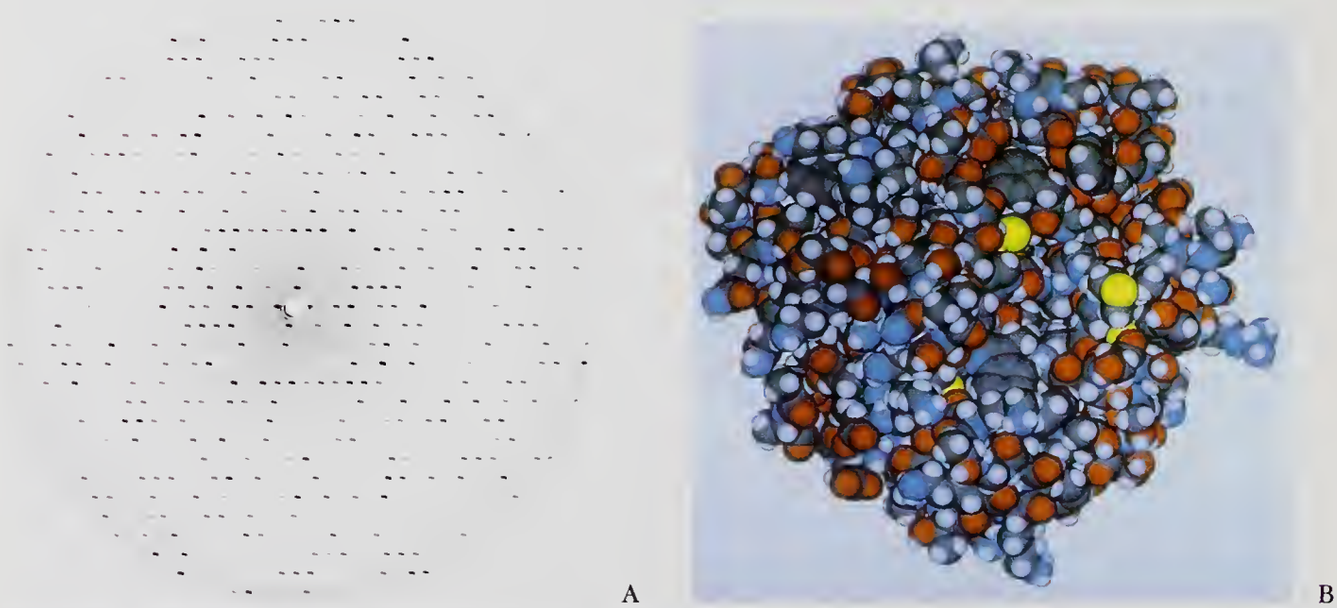


Figure 26. A: The x-ray diffraction pattern given by the enzyme chymotrypsin.

Courtesy of Thomas A. Steitz.

B: A representation of the three-dimensional structure of chymotrypsin. Carbon atoms are shown in black, nitrogen in blue, oxygen in red, hydrogen in white, and sulfur in yellow. The diameter of the enzyme is about 45 Å (somewhat less than a millionth of an inch). The hydrophilic side chain of arginine-145 is clearly visible projecting outward from the right side of the molecule. The ridges and grooves on the surface of the chymotrypsin molecule are as unique as the mountains and craters of the moon, and herein lies the fulfillment of the lock-and-key mechanism hypothesized by Emil Fischer at the turn of the century.

Courtesy of Molecular Simulations, Waltham, MA.

2. *Time-resolved images of events at the active site of an enzyme.* New ways of using synchrotron x-ray radiation permit very rapid measurements of diffraction data, so that in principle it should be possible to follow the structural changes that occur during an enzymatic reaction. If we understood these changes, it might be possible to develop enzymes with usefully altered properties and to synthesize enzyme inhibitors with enhanced specificity.

3. *The role of molecular recognition in the regulation of cellular activity.* How do proteins that control transcription recognize specific DNA sequences? How do cell surface proteins in the immune system recognize and present antigens? The answers to these questions are beginning to emerge from crystal structures of molecular complexes, such as the complexes formed by regulatory proteins with DNA, those formed by binding proteins with their appropriate ligands, of antibodies with antigens, and of MHC molecules with peptides. These current efforts give considerable promise for understanding how hormones or neurotransmitters trigger a cascade of events that involves the formation and dissociation of protein assemblies inside cells. A large and medically significant class of regulatory interactions involves the protein products of oncogenes or proto-oncogenes.

As our knowledge of important proteins rapidly increases, we can only hope that our capacity to anticipate aspects of structures not yet determined will keep pace. The goal of accurately predicting the three-dimensional structure of any protein from its amino acid sequence is still a long way off. But recent advances in computational chemistry make it possible to predict the effects of small perturbations, such as point mutations, on the folding of a protein and to calculate differences in binding free energies for related ligands. And systematic approaches to designing drugs, such as antagonists or inhibitors of enzymes, are beginning to emerge now that we can carry out meaningful calculations on known structures.

Nuclear magnetic resonance (**NMR**) methods offer an alternative route to determining the three-dimensional structures of peptides and small proteins. The DNA-binding domains of transcriptional activators and repressors are good candidates for this type of analysis. The past two years have seen the determination by NMR of the structure of a developmentally important DNA sequence known as a **homeodomain** and of a DNA-binding structure known as a **zinc finger**. NMR has the great advantage that it circumvents the need to crystallize the protein to be studied. At the other end of the size scale, imaginative combinations of light and electron microscopy have begun to reveal important patterns and regularities in very large structures, such as chromosomes, viruses, and receptors. New methods for recording images and enhancing contrast in light microscopy make it possible to record in real time the events of intracellular transport or the process of chromosome condensation. As the molecules that generate these large-scale intracellular motions are characterized, it should become possible to relate such changes to the specific molecular recognition events that control them. It is fortunate, but not coincidental, that as biologists have become increasingly aware of the need to know the precise structure of the molecules that mediate the phenomena they are interested in, a whole range of new experimental methods has been developed, and a new generation of structural biologists has emerged to assist them and to advance their understanding of the complex relationships between structure and function.

Investigators in the Structural Biology Program

Agard, David A., Ph.D.
Brünger, Axel T., Ph.D.
Burley, Stephen K., M.D., D.Phil.
Deisenhofer, Johann, Ph.D.
Fox, Robert O., Ph.D.
Harrison, Stephen C., Ph.D.
Hendrickson, Wayne A., Ph.D.
Kuriyan, John, Ph.D.
Matthews, Brian W., Ph.D., D.Sc.
Pabo, Carl O., Ph.D.
Quiocho, Florante A., Ph.D.
Sedat, John W., Ph.D.
Sigler, Paul B., M.D., Ph.D.
Sprang, Stephen R., Ph.D.
Steitz, Thomas A., Ph.D.
Wiley, Don C., Ph.D.

W. Maxwell Cowan, M.D., Ph.D.
Vice President and Chief Scientific Officer

Electrical Activity of Nerve Cells

Paul R. Adams, Ph.D.—Investigator

Dr. Adams is also Professor of Neurobiology and Behavior and of Pharmacological Sciences at the State University of New York at Stony Brook. He received his B.A. degree in physiology and pharmacology from Cambridge University and his Ph.D. degree in pharmacology from the University of London. His postdoctoral work was done with Bert Sakmann at the Max Planck Institute, Göttingen, and with Philippe Ascher at the Ecole Normale, Paris. Dr. Adams is currently a MacArthur Fellow. He was elected Fellow of the Royal Society.

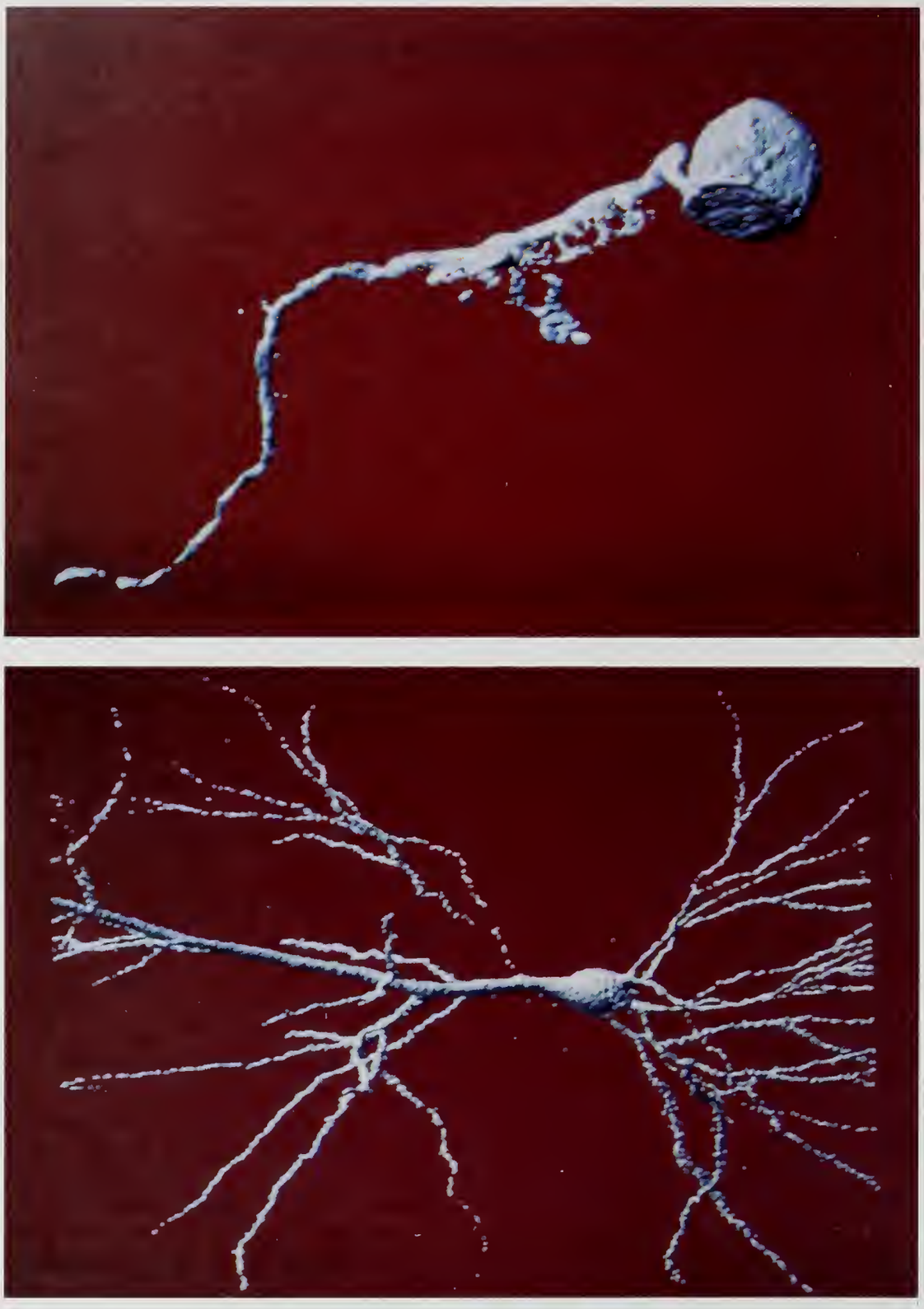
NERVE cells are specialized to generate, transmit, and receive rapid electrical messages. Electrical impulses, called action potentials, last about 1,000th of a second and can travel along specialized nerve cell extensions at speeds over 100 mph. Chemical transmitter substances released onto the nerve cell by other nerve cells control the precise timing of these electrical pulses. We are trying to understand how these pulses are generated and how transmitters impinging on the cell control them.

Cell membranes are normally effective barriers to the movement of ions (electrically charged atoms) between the cell environment and the cell interior. This insulating property allows the inside of a nerve cell to have a different electrical voltage from the outside or from a neighboring cell. The electrical activities described above are regulated by special protein molecules, called ion channels, which are embedded in the cell membrane. There are many types of ion channel. Each type has a specific role, but all have in common a unique feature that allows certain ions to travel easily through them. The protein chains that make up an ion channel molecule are arranged to create a minute tunnel, through which certain types of ions—for example, sodium, potassium, or calcium ions—can quickly move.

The direction that the ion moves is not controlled by the tunnel but by the ion concentrations and the transmembrane voltage. The tunnel does, however, control the type of ion that moves. Thus the sodium channel only allows sodium ions to pass. Because sodium ions are abundant outside, but not inside, nerve cells, the existence of open sodium channels leads to an inward stream of sodium ions, making the cell interior positive. On the other hand, when potassium channels open, potassium streams *out* of the cell, making it negative. Because these tunnels are not always open (indeed are closed most of the time), it is supposed that the channel must have some sort of gate.

There are many types of potassium channel, differing according to their molecular structure (see the articles in this volume by Richard Aldrich and Lily Jan), speed of opening and closing, and the way in which their gates are controlled. We are studying potassium channels in bullfrog sympathetic ganglion cells to understand better how they are controlled and how they contribute to the electrical activity of nerve cells. We have been particularly intrigued by a channel we call the M channel. This channel is turned off as a result of the binding of certain neurotransmitters (chemicals released from nerve endings) to receptors on the cell surface. As a result, less potassium leaves the cell, which is therefore less negative and more able to fire electrical pulses. Turnoff of this channel occurs via activation of a G protein (see the article by John Exton). However, after the neurotransmitter has been removed, M channels turn back on and transiently become more numerous than initially observed. We have shown that calcium and arachidonic acid are involved in this overshooting response. Various levels of calcium were perfused into nerve cells while they were visualized with a special calcium-detecting microscope. Neurotransmitter stimulation of receptors produces a small calcium signal that is sufficient to increase the activity of M channels. However, this is not seen until the concomitant G protein–mediated suppression of the channels is terminated by removing the transmitter.

M channels work in concert with many other types of channels, which we have also characterized. This information can then be combined with studies of calcium diffusion and membrane geometry to predict completely the cell's electrical output. We are making detailed quantitative morphological measurements of cells that have been previously characterized electrophysiologically in both bullfrog ganglia and mammalian hippocampus and lateral geniculate. This is achieved by automatic three-dimensional reconstruction of dye-filled cells, which are optically sectioned using confocal microscopy. Our voxel-based reconstructions, developed in collaboration with the computer science department at Stony Brook, can then be used as a platform for Monte Carlo simulations of the movements of single ions in small cell structures, such as dendritic spines.



Three-dimensional computer reconstructions of nerve cells. The upper image shows a sympathetic ganglion cell of a bullfrog; the lower image, a hippocampal pyramidal cell of a rat.

Research of Rick Avila, Barry Burbach, Jim Monckton, Ted Carnevale, Arie Kaufman, and Paul Adams.

Three-Dimensional Macromolecular and Cellular Structure

David A. Agard, Ph.D.—Associate Investigator

Dr. Agard is also Associate Professor of Biochemistry and Biophysics at the University of California, San Francisco. He did his undergraduate work at Yale University with Frederic Richards, Hal Wyckoff, and Thomas Steitz. He received his Ph.D. degree in chemical biology from the California Institute of Technology, where he studied with Robert Stroud and began a continuing collaboration with John Sedat. His postdoctoral work was done on high-resolution electron microscopic crystallography at the MRC Laboratory of Molecular Biology in Cambridge, England, with Richard Henderson. There he also began the cloning of the α-lytic protease gene with Sydney Brenner.

WE study chromosome structure in close collaboration with John Sedat (HHMI, University of California, San Francisco); hence only a subset of these studies will be discussed here. Our primary aim in this area is to provide a physical basis for understanding chromosome behavior and function by directly determining the three-dimensional structure of chromosomes as a function of both transcriptional state and the cell cycle stage. Current efforts are aimed at determining how fibers of nucleosomes are folded into higher-order structures within the chromosome and what role specific chromosomal proteins play in determining these structures.

We are using intermediate voltage electron microscope (IVEM) tomography to examine higher-order chromosome structure. In the past year significant improvements in the quality of the three-dimensional reconstructions have led to the first new insights into the structure of the 30-nm fiber (in collaboration with Chris Woodcock, University of Massachusetts). We are now beginning to be able to trace the paths of the 30-nm chromatin fibers within telophase chromosomes. It is clear that the existing models of chromatin structure are seriously flawed. In addition, we have made dramatic steps toward our goal of fully automating the complex task of collecting three-dimensional IVEM tomography data. This will simplify the arduous task of collecting 100–150 images and should dramatically reduce radiation exposure (from 3 hours to about 5 minutes). Significant progress has also been made on understanding the mechanism of image formation for thick specimens and on applying a powerful new approach to the problem of electron microscopic (EM) reconstruction. Together with a new staining approach and cryopreparation methods, these improvements should allow us to trace the paths of the 30-nm fibers throughout the chromosome and will finally make this exciting structural approach available to the general cell biologist.

Structural Basis of Enzyme Specificity

By combining solution kinetic analysis, x-ray crystallographic structural analysis, and site-directed mutagenesis, we have been able to probe the structural basis for substrate specificity in unprecedented detail, using α-lytic protease as a model system. The availability of peptide boronic acid inhibitors, which act as excellent mimics of the reaction transition state, or nearby intermediates, has allowed us to use x-ray crystallography to examine the complex set of interactions between enzyme and substrate that mediate specificity.

By mutation, we have been able to alter dramatically the pattern of substrate specificity while maintaining or even increasing enzyme activity. Approximately 40 high-resolution, extremely well-refined crystal structures have now been determined and analyzed. Detailed structural analyses of three mutants as free enzymes and as complexes have provided surprising insights into the mechanism of specificity and have indicated the crucial role that protein flexibility plays in selectivity. During the past year we have made numerous other mutations and examined their kinetic and structural properties. Of particular note is a mutation that alters specificity indirectly by changing active-site flexibility. This remarkable finding is the first demonstration that residues beyond those that directly contact the substrate can play a significant role in determining the patterns of specificity.

Not long ago we began a collaboration with Vladimir Basus (University of California, San Francisco) to perform a complete analysis of α-lytic protease structure in solution by nuclear magnetic resonance (NMR) methods. We have now made ^{15}N,^{13}C doubly labeled enzyme and have collected sufficient high-resolution three-dimensional NMR data sets to permit complete backbone assignment—a significant task for such a large protein. Currently we are working on the assignments and on the collection of data for the complete side chain assignments. We anticipate that the NMR data will provide insights into correlated motions within the enzyme and be crucial for the folding studies described below.

Last year we had developed a new method for predicting the energetics of protein-substrate interactions. This approach (based on Ponder-Richards rotamers combined with energetics and solvation terms) can predict k_{cat}/K_m with stunning accuracy. We have used this method to design a new enzyme with particular properties, and so far the results have been remarkable. This year we have been working to extend these results to other systems and to improve their accuracy. We have been able to predict quantitatively 100 values of k_{cat}/K_m for subtilisin with equivalent accuracy. This approach is being expanded to allow its use in drug design situations and for modeling proteins based on the structure of a homologous protein.

Structural and Biochemical Probes of Folding of α-Lytic Protease

α-Lytic protease is synthesized as a preproenzyme. Experiments in *Escherichia coli* have demonstrated that the 166–amino acid precursor domain is required for the proper folding of the 198–amino acid protease domain. Furthermore, we have shown that the pro region does not have to be attached covalently in order to function. The development of the *in vitro* folding system has led to the unprecedented ability to trap and purify a stable folding intermediate under nondenaturing conditions. Although the intermediate is unable to fold by itself, it is rapidly converted to active enzyme upon addition of the pro region. Analysis of the folding kinetics proves that the pro region acts essentially as a "foldase" to stabilize the transition state for folding, thereby speeding up folding by more than 10^7.

Current efforts are focused on characterizing the folding intermediate and analyzing the folding reaction. We plan to use a combination of approaches—including NMR and crystallography—to probe the structure of the intermediate and the role of the pro region in the final stage of folding.

Receptor-Ligand Targeting and Human Cholesterol Metabolism

Apolipoprotein E is an important protein in cholesterol metabolism in mammals. Responsible for targeting chylomicrons, very low density lipoprotein (VLDL), and high-density lipoprotein (HDL) particles to the low-density lipoprotein (LDL) receptor, apolipoprotein E has a major role in triglyceride and cholesterol metabolism. The protein itself has two distinct structural and functional domains: the amino-terminal 22-kDa domain contains the receptor-binding functionality; lipid binding resides primarily with the 10-kDa carboxyl-terminal domain. In collaboration with the Mahley group at the Gladstone Foundation Laboratories for Cardiovascular Disease, we obtained crystals of the 22-kDa receptor-binding domain. Last year we reported that the protein is an unusually elongated four-helix bundle. Although the surface is exceptionally charged, there is a precise balance between groups with a positive charge and groups with a negative charge, except in what we believe is the receptor-binding region. We have recently finished the analysis of a naturally occurring human mutant that significantly reduces receptor binding and can lead to premature atherosclerosis. The structure reveals that the mutant disrupts the complex set of surface salt bridges, which then causes a key arginine required for binding to be recruited into a salt bridge. Current efforts are aimed at generating a soluble fragment of the LDL receptor-binding domain for crystallographic study.

Molecular Mechanisms of Ion Channel Function

Richard W. Aldrich, Ph.D.—Associate Investigator

Dr. Aldrich is also Associate Professor of Molecular and Cellular Physiology at Stanford University. He received his B.S. degree in biology from the University of Arizona and his Ph.D. degree in neuroscience from Stanford. He carried out postdoctoral research at Yale University with Knox Chandler and Charles Stevens. Before returning to Stanford, he was Assistant Professor of Molecular Neurobiology at Yale. Among his awards are a Searle Scholars Award, the Young Investigator Award of the Society for Neuroscience, and the Young Investigator Award of the Biophysical Society.

ION channels are the molecular units of electrical signaling in cells. They are proteins that regulate the movement of ions, such as sodium, calcium, and potassium, into and out of cells. They are responsible for the conversion of external sensory signals to the electrical language of the nervous system and for the integration of these signals to generate appropriate behavior. In addition, ion channels are important for the generation and regulation of the heartbeat, for contraction of muscles, and for the release of hormones in the bloodstream. A very large variety of ion channel types are found in the body. They are specialized to select for certain species of ions and to open and close in response to a number of different stimuli, such as the binding of a neurotransmitter molecule or a change in the voltage that exists across a cell's membrane. Our laboratory is interested in the molecular mechanisms of ion channel function. One of our major goals is to understand the conformational changes that occur as the channels open and close in response to appropriate stimuli.

Voltage-gated ion channels are an important functional class. As their name implies, they can open in response to changes in the electrical potential across the cell membrane, a property crucial for the generation of electrical signals and for the transmission of information throughout the body. These molecules have a way to measure the electrical potential and open accordingly. In addition, a number of these channels inactivate or become unavailable for opening after use. In recent years we have studied the molecular mechanisms of inactivation of a class of potassium channels that were cloned in *Drosophila*. These channels are products of the *Shaker* gene. They exhibit the fastest inactivation of any potassium channels yet cloned. William Zagotta and I began by using single-channel recording methods to study the gating properties of wild-type *Shaker* channels in their native tissue. Such methods allow us to record the behavior of a single-channel molecule as it opens and closes on a millisecond time scale.

We determined that the conformational changes associated with opening the channel depended strongly on the membrane voltage and therefore involved a substantial rearrangement of an electrically charged part of the channel in the membrane. On the other hand, the inactivation process did not involve significant charge rearrangement. This result, combined with our ability to alter inactivation by internal enzymes, led us to the hypothesis that inactivation involved a conformational change on the inside of the membrane that blocked the flow of potassium ions through the channel.

Dr. Zagotta, Toshinori Hoshi, and I further studied this hypothesis by making altered channels with recombinant DNA methods and expressing the normal and altered channels in frog oocytes. Our results are strikingly consistent with the "ball and chain" model of inactivation originally proposed for the voltage-dependent sodium channel by Armstrong and Bezanilla in 1977. The following model of the molecular mechanism of inactivation emerges from our results. The amino terminus of the *Shaker* channel acts as an inactivation particle or internal plug for the channel. When the inactivation particle is bound to the receptor, the channel closes. We tested this model further by applying a solution containing free synthetic inactivation particle to the inside face of mutant channels that did not inactivate. In the presence of the synthetic inactivation particle, the mutant channels regained inactivation, consistent with the ball and chain mechanism.

Ruth Murrell-Lagnado and I have used synthetic peptide inactivation particles with altered amino acid composition to determine the important features that influence an ability to bind to the internal mouth of the channel. The naturally occurring amino acid sequence can be divided into an uncharged region and a highly charged region. The charged amino acids interact with negative charges at or near the mouth of the channel to influence the rate of binding. The more positive charges in the charged region, the faster the binding rate. Negative charges in the charged region inhibit binding. A surprising result is that the net charge in this region seems to be the important factor for the binding rate, regardless of

which of the amino acids are charged. Conversely, the uncharged region is important for determining the unbinding rate. This work was supported in part by the National Institutes of Health and the National Institute of Mental Health.

Shaker potassium channels also exhibit a slower inactivation process, which can be seen both in wild-type channels and after the faster inactivation process has been removed by mutagenesis. The slow inactivation does not require intact fast inactivation. Although it also does not involve rearrangement of charge in the membrane, this slower inactivation seems to involve a mechanism different from the fast process. In collaboration with Kathleen Choi and Gary Yellen, we have found that the slow inactivation is affected by external agents, suggesting that the conformational changes for this process involve external structures. Dr. Hoshi, Jose Lopez Barneo, and I have found that the rate of this type of inactivation is profoundly affected by external ions. Normally sodium is the predominant positively charged ion outside the cell. When a small amount of potassium is added to the external solution, as could occur during epileptic activity in the brain or other pathological states, the inactivation is significantly slowed. We have studied the dependence of the inactivation rate on the type of external ion present and have determined that the ability of an ion species to slow inactivation correlates with its permeability in the channel. Ions that get into the channel better from the outside inhibit inactivation more effectively. These results are consistent with a hypothesis whereby a channel that is occupied by an ion cannot undergo slow inactivation.

The slow inactivation process occurs by greatly different rates in variants of the *Shaker* channel with differences in structure at the carboxyl end of the protein. We have made mutations in both of these variants and have localized the difference responsible for the differences in slow inactivation to a single hydrophobic amino acid in a membrane-spanning region of the channel molecule. Other amino acid substitutions at this position have dramatic effects on gating, with larger hydrophobic amino acids leading to slower inactivation.

Divergent Members of the SRY Family of Transcriptional Regulators Bind an Insulin-Responsive Element, IRE-A

Maria C. Alexander-Bridges, M.D., Ph.D.—Assistant Investigator

Dr. Alexander-Bridges is also Assistant Professor of Medicine at Harvard Medical School and Assistant in Medicine at Massachusetts General Hospital, Boston. She received her M.D. and Ph.D. degrees from Harvard University Medical School, where she was a member of the Harvard-MIT Health Sciences and Technology program, which is geared toward students interested in academic medicine. She developed an abiding interest in hormonal regulation of cellular metabolism and, as a graduate student in physiology, investigated the mechanism of insulin-stimulated phosphorylation of cellular proteins. Dr. Alexander-Bridges then served as an intern and resident at the Johns Hopkins University. After subspecialty training in endocrinology at Massachusetts General Hospital, she was a postdoctoral fellow with Howard Goodman.

THE main function of insulin is to allow a starved animal to adapt to a glucose load. Insulin does this by activating enzymes that promote energy storage and inactivating enzymes that break down energy stores. Over the long term, insulin alters the expression of these enzymes by regulating the transcription of their genes. The work in our laboratory is directed toward understanding the mechanism of insulin action on expression of metabolically active genes in tissues that modulate glucose utilization. We use the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene as a model gene for the anabolic effects of insulin, because it is highly regulated by insulin 1) in cultured 3T3 adipocytes, 2) during nutritional manipulations such as fasting and refeeding, and 3) during the induction and treatment of diabetes.

Transgenic animals that express GAPDH–growth hormone fusion genes have been made in collaboration with Jeung Yun and Tom Wagner (Ohio State University). Using these animals, we have confirmed that GAPDH gene transcription is regulated *in vivo* by nutritional manipulations that lead to hyperinsulinemia. For example, fasting a rat and refeeding it a high-carbohydrate, low-fat diet increases circulating glucose and insulin levels, resulting in the induction of glycolytic and lipogenic enzymes and the repression of gluconeogenic and lipolytic enzymes.

We have identified a cis-acting sequence, insulin-responsive element A (IRE-A), in the upstream region of the GAPDH gene that interacts with an insulin-responsive DNA-binding protein (IRP-A). Activation of GAPDH gene expression in insulin-responsive tissues correlates with the presence of IRP-A. Within one hour of exposure of 3T3 adipocytes or H35 hepatoma cells to insulin, the activity of this protein is increased two- to fourfold. The activity of IRP-A is induced four- to eightfold in liver and fat during the process of refeeding a fasted rat a high-carbohydrate, low-fat diet. These observations support the importance of GAPDH gene regulation. In muscle, where GAPDH activity is not rate limiting and is not regulated by insulin, IRP-A binding is not detectable.

The wild-type IRE-A motif was used to isolate a clone from a rat adipocyte library, using the Singh-Sharp Southwestern screening approach. The cloned cDNA encodes a protein (IRE-ABP) that binds IRE-A DNA with sequence specificity that overlaps that of the adipocyte IRP-A nuclear extract protein. IRE-ABP is expressed in liver and fat but not in muscle, which provides an explanation for the tissue-specific regulation of GAPDH gene expression. Expression of IRP-A mRNA is induced during the process of fasting and refeeding. In contrast, one hour of insulin exposure of cells does not appear to alter expression of the IRP-A mRNA. IRE-ABP footprints the upstream region of genes that are inhibited and genes that are stimulated by insulin. We presume that IRE-ABP will regulate the transcription of metabolic genes with diverse functions to change the phenotype of the fat and liver cell during the switch from the fasted to the refed state that is initiated by glucose and insulin. Studies are in progress to determine whether IRE-ABP mediates the effect of insulin alone or in association with another protein.

IRE-A DNA-binding Protein Shares Binding Specificity with the Testis-determining Factor

Surprisingly, the HMG (high-mobility group) box domain of IRE-ABP is 68 percent identical to *SRY*, the testis-determining factor, and is 98 percent identical to an autosomal gene that was isolated during the process of screening a whole mouse embryo cDNA library for *SRY*-related sequences. Furthermore, IRE-ABP and *SRY* share DNA-binding specificity for IRE-A. Although IRE-ABP shows markedly higher affinity for the IRE-A motif, the nucleotides protected by these two divergent family members are essentially identical. The sequence in IRE-A that is contacted by IRE-ABP and *SRY* is highly conserved: the sequence 5′-Py-ctttg(a/t)-3′, previously defined by Katherine Jones and her colleagues as a consensus motif,

is contained in several T cell–specific genes that are bound with high affinity by TCF-1α (T cell factor 1α). Thus diverse members of the HMG family of proteins may modulate transcription through a similar spectrum of sequences that contain a core motif. Identification of such a motif may provide a clue to the identity of important physiological targets of the IRE-ABP and *SRY*-like family of transcriptional regulators.

The laboratories of Peter Goodfellow and Robin Lovell-Badge have identified *SRY* as the testis-determining region on the basis of genetic evidence. Several patients with an XY genotype who failed to differentiate to the male phenotype have mutations in the HMG box domain of *SRY*. The *SRY* locus is widely presumed to encode a sequence-specific DNA-binding protein because it contains an HMG domain within its open reading frame. The DNA-binding properties and mechanism by which this protein regulates transcription have not been defined. Certain of the mutations that were found in the sex-reversed patients were not *de novo* mutations, and thus it was not possible to deduce whether these mutations were completely unrelated to the sex-reversed phenotype or contributed to it. Because IRE-ABP was isolated on the basis of its ability to bind the IRE-A motif, it was possible to examine the effect of these mutations on the binding affinity of IRE-ABP and *SRY* derivatives. Derivatives that contained the mutations associated with sex reversal at positions 3 and 7 in the HMG box domain showed marked impairment in their ability to protect the IRE-A motif from DNase I digestion. Furthermore, derivatives of IRE-ABP and *SRY* that contain a switch between IRE-ABP and *SRY* at position 3 in the HMG box domain show altered binding affinity for the IRE-A motif. Thus positions 3 and 7 appear to be important determinants of binding affinity for this family.

Future Directions

Identification of the IRE-A binding protein as a member of the *SRY* family suggests many avenues of investigation. Studies are under way to define the binding specificity of IRE-ABP– and *SRY*-like family members. Definition of the preferred binding site for these related proteins will facilitate the identification of other insulin-sensitive genes that are regulated by IRE-ABP and potential targets of *SRY*. For example, insulin simultaneously activates and inhibits diverse metabolic processes to alter the flux of metabolites into glycogen and fat. We have located the proposed consensus sequence in the upstream region of genes that are regulated in a positive and negative direction by insulin, and we can now establish whether IRE-ABP plays a role in regulating these diverse metabolic processes.

The observation that the *SRY* protein and IRE-ABP share binding specificity for a sequence located in a glycolytic gene implies that these proteins carry out similar functions in the specific tissues in which they are expressed. It is clear, for instance, that spermatogenesis in the adult testis requires high lactate production; thus both gene products may be involved in regulating glycolysis/gluconeogenesis in their respective target tissues. Conversely, IRE-ABP–like genes may play a role in regulating processes that show sexual dimorphism in adult tissues or promoting differentiation of its target tissues during embryogenesis.

Studies on the regulation of IRP-A gene expression have led to an understanding of the mechanism by which insulin modulates the expression of specific genes in specific tissues involved in the maintenance of normal glucose and lipid metabolism. Ultimately we expect these studies will lead to an understanding of the signal transduction process by which insulin modulates the expression of these genes. Understanding the hormonal control of lipid metabolism at a molecular level will provide insights into two disease states of major importance, obesity and diabetes.

Genetic Mechanisms Involved in the Generation of the Antibody Repertoire

Frederick W. Alt, Ph.D.—Investigator

Dr. Alt is also Professor of Genetics and Pediatrics at the Children's Hospital, Boston, and Harvard Medical School. He obtained his undergraduate degree in biology from Brandeis University and his Ph.D. degree in biological sciences from Stanford University, where he worked with Robert Schimke. He did postdoctoral work with David Baltimore at the Massachusetts Institute of Technology, after which he was Professor of Biochemistry and Microbiology at Columbia University College of Physicians and Surgeons. His honors include the Irma T. Hirschl Career Scientist Award, the Searle Scholars Award, and the Mallinckrodt Scholar Award.

WE are interested in the molecular mechanisms that underlie the generation of a specific immune response. The mammalian immune system functions through complex interactions between various cells and their products. Cells that effect specific immunological responses fall into two general categories: B lymphocytes that mediate humoral immunity (i.e., production of antibodies against foreign antigens) and T lymphocytes that mediate cellular immunity (e.g., foreign graft rejection). During the earlier stages of lymphocyte development, stem cells proceed through a developmental program that ultimately leads to the generation of a multitude of individual B or T lymphocyte clones (each clone is an essentially identical set of cells derived from a common parent). Each set of clonal cells expresses a novel receptor on its surface that will recognize a unique set of antigens.

The ability of the immune system to respond specifically to a vast array of antigens results in substantial part from the unique organization of the genes that encode antigen receptor proteins. Unlike most genes, antigen receptor genes are not inherited intact from our parents. Instead, these genes are encoded in cassettes (gene segments) in the germline and are assembled into complete genes only during the somatic differentiation of lymphocytes. This lymphocyte-specific gene assembly process is referred to as VDJ recombination. Because there are many individual cassettes that encode various portions of antibodies and because these can be put together in various combinations or in various ways, the body can randomly assemble a vast array of different antibody genes from a limited amount of genetic material.

Much of our work is aimed at determining the genetic mechanisms by which antibody genes are assembled from gene segments, the role of the gene assembly process in the generation of antibody diversity, and the mechanisms that regulate this gene assembly process and ensure that it occurs only in appropriate cell types. We are also working on the elucidation of molecular signals that control the various steps of B lymphocyte differentiation.

The enzymatic system involved in the assembly of antigen receptor genes involves a variety of different activities. Some of these activities are expressed specifically in developing lymphocytes and are likely involved in the early processes of gene assembly, including recognition of the specific gene segments that will be joined and cutting them away from the surrounding genetic material. Other activities employed in VDJ recombination are likely expressed in many cell types where they may be involved in other processes, such as replication of genetic material or repair of lesions in genetic material (DNA repair). The more widely expressed activities are probably recruited by the lymphocyte-specific components of the system to carry out certain aspects of the joining event, such as pasting together the "cut" gene segments.

The lymphocyte-specific components of the VDJ recombination system are likely encoded by two genes (recently isolated by David Baltimore and David Schatz) referred to as recombination-activating genes 1 and 2 (*RAG-1* and *-2*). Simultaneous expression of these genes was found to occur only in developing lymphocytes; artificially induced expression of the two genes (simultaneously) in nonlymphoid cells confers them with the ability to undergo VDJ recombination. However, these genes have been found to be expressed in other tissues, including the central nervous system, fueling speculation that VDJ recombination activity or the activity of one or the other of the two *RAG* gene products might be involved in developmental processes in tissues other than the immune system.

To test unequivocally the function of the *RAG-2* gene, we have used gene-targeting technology to eliminate a copy of this gene in mouse embryonic stem cell lines. These cells were then introduced into developing mouse embryos, where they were incorporated into the germ cells of the resulting chimeric mouse. Mice that carried a single copy of the targeted mutation in their germline were interbred to generate animals that completely lack the gene. Mice that lack the *RAG-2*

gene are viable and appear normal at birth. However, they routinely develop severe infections in the first months of life due to the total absence of mature T or B lymphocytes—a severe combined immune deficiency. These animals do have large populations of very immature B and T lymphocytes in their primary lymphocyte differentiation organs; however, these cells are unable to initiate the VDJ recombination process. Because the *RAG-2*–deficient mice have no defects in any cell type besides lymphocytes, we have concluded that VDJ recombination activity and *RAG-2* gene function are required only for lymphocyte development and not for any other developmental process.

Our group has also helped characterize the basis of a naturally occurring murine mutation that when homozygous leads to severe combined immune deficiency (the *scid* mutation). Like *RAG-2*–deficient mice, animals homozygous for the *scid* mutation lack functional B or T lymphocytes due to inability to assemble antigen receptor gene segments correctly. However, mice homozygous for the *scid* mutation produce normal *RAG-1* and *-2* gene products and can efficiently initiate the VDJ recombination process. Completion of the joining process is blocked in *scid* mutant cells. Moreover, both lymphoid and nonlymphoid cells of mice homozygous for the *scid* defect are unable to repair certain types of lesions in their genetic material. Thus the product of the gene affected by the *scid* mutation appears to represent an example of a component of the VDJ recombination system that is used both in VDJ recombination and in more general cellular processes.

To define further generally expressed genes potentially involved in VDJ recombination, we provided the lymphocyte-specific components of this system (*RAG-1* and *-2*) to mutant Chinese hamster ovary cell lines that are defective in ability to repair breaks in their DNA. By introducing antigen receptor gene segments into these lines, we could ask whether the cells were capable of performing correct VDJ recombination. By this approach, we have now derived a series of independent genetic mutations that affect both DNA repair and VDJ recombination. Introduction of specific human chromosomes into the mutant cell lines restored both ability to repair DNA and to undergo VDJ recombination. We are now in the process of isolating the defective genes in these cells. Elucidation of these genes will provide insight into the VDJ recombination process and may also provide information relevant to understanding the basis for several human diseases that affect both ability to repair genetic lesions and to generate a normal immune system.

We are also analyzing several other novel mouse models generated to study various factors involved in lymphocyte differentiation and the generation of the immune response. One such model is a mutant mouse line that cannot produce endogenous antibody molecules because it has been genetically engineered to lack germline gene segments necessary for forming functional antibody genes. We are using these animals to study the role of antibody gene products in regulating antibody gene assembly and lymphocyte development. A potential practical use of such mice may be achieved by breeding them with transgenic mouse lines that have been engineered to contain functional human antibody gene cassettes. The hope is that the hybrid animal will rely on the human antibody genes for its immune system, providing a more effective method of generating tailor-made human antibodies.

Control of Cell Fate During Vertebrate Neuronal Development

David J. Anderson, Ph.D.—Assistant Investigator

Dr. Anderson is also Associate Professor of Biology at the California Institute of Technology and Adjunct Assistant Professor of Anatomy and Cell Biology at the University of Southern California School of Medicine. He received his A.B. degree in biochemical sciences from Harvard College and a Ph.D. degree in cell biology from the Rockefeller University. He then did postdoctoral research in molecular neurobiology at Columbia University. He is the recipient of an NSF Presidential Young Investigator Award, a Sloan Foundation Fellowship in Neuroscience, and the 1989 Charles Judson Herrick Award in Neurobiology from the American Association of Anatomy.

WE are interested in how different types of nerve cells are generated during the development of the vertebrate nervous system. We have chosen to work on the peripheral autonomic nervous system, which is simpler and more accessible than the brain. Our studies have focused on the development of two specific cell types: the sympathetic neurons that lie in a chain of ganglia along the spinal cord and the chromaffin cells of the adrenal medulla. These two cells are closely related, yet distinct in major respects. The former are true neurons, with long branching axons and dendrites that send and receive electrical signals. The latter are small, round secretory cells that release epinephrine (adrenaline) into the bloodstream during fright or excitement.

Studies in a number of laboratories, including our own, have established that these two cell types derive from a common progenitor cell. This cell arises on top of the neural tube (the developing spinal cord), as part of a transient structure called the neural crest. Like parachutists, the neural crest cells peel off the neural tube in a wave and migrate downward through the embryo. Some of them arrest their migration in a chain of small clumps along a blood vessel, where they eventually become sympathetic neurons. Others continue their migration downward to invade the developing adrenal gland, where they become chromaffin cells.

Using a sophisticated fluorescence-activated cell sorter and specific monoclonal antibodies to tag the cells, we have succeeded in isolating chromaffin cell precursors from the fetal adrenal gland of the rat. By manipulating the cell culture environment, we have shown that these precursors have two possible developmental fates: if glucocorticoid hormones are added to the growth medium, mimicking the environment of the adrenal gland, then the precursors develop into chromaffin cells. If, on the other hand, fibroblast growth factor (FGF) and nerve growth factor (NGF) are added to the medium, the precursors develop into sympathetic neurons. This suggests that the fate of these cells is determined in large part by signals in the environments to which they migrate. However, these precursors seem to have lost the ability to give rise to other derivatives of the neural crest, such as glial cells. Therefore, while these precursor cells have a choice of fate, it is a restricted one.

Control of Neuronal Survival by Neurotrophic Factors

One problem in studying the chromaffin-neuron precursor cell at the molecular level is the small number of cells that can be recovered after isolation from rat fetuses. To circumvent this problem, we have applied recently developed techniques to immortalize the cells. Using a defective retrovirus as a "disposable molecular syringe," we have injected the cells with a gene, v-*myc*, that allows them to divide forever in the culture dish. In this way we can generate an endless supply of cells that can be used for experiments at any time, without the need to perform long hours of dissection. Fortunately, these immortalized precursor cell lines still appear capable of undergoing differentiation into sympathetic neurons when exposed to FGF and NGF.

Sympathetic neurons, like other neurons, depend on specific neurotrophic factors for their survival. Neurotrophic factors are proteins secreted by the tissues innervated by the neurons; these proteins nourish the neurons and keep them alive. Different types of neurons use different neurotrophic factors: NGF is the factor for sympathetic neurons. The specificity of NGF action is due to a specific receptor protein present in the membrane of the nerve cell. This receptor binds NGF and sends a signal to the cell, keeping it alive. We are using the immortalized precursor cell line to study how developing neurons acquire NGF receptors during development. The NGF receptor has two components (subunits); our data suggest that different factors induce the expression of these two subunits. One factor that seems to be important for the induction of functional NGF receptors is electrical activity. This suggests that the stimulation of the developing

neuron by its presynaptic partner may endow it with the machinery to respond to key survival factors secreted by its postsynaptic target. In this way, the establishment of functional connections between a neuron, its input, and its output target could be coordinated.

Control of Neuron-Specific Gene Expression

As it differentiates, the chromaffin-neuron precursor expresses genes that are not expressed in other cell types. How such specific expression is achieved is a basic question in modern developmental biology. Studies in blood and lymphoid cells have revealed that genes such as that encoding hemoglobin are specifically turned on in the appropriate cells by activator proteins, present only in those cells. We have studied a neuron-specific gene, SCG10, and found that a different mechanism restricts its expression to developing neurons. SCG10 appears to be specifically turned off in all tissues except neurons; this repression is somehow relieved in neurons and not in other cells. We have made advances in identifying a protein involved in the repression of SCG10. This "silencer" protein is present in nonneuronal cells and tissues but not in neuronal cells, consistent with the idea that it shuts off SCG10 everywhere except in the nervous system. This same silencer protein also appears to be involved in shutting off other neuron-specific genes. This suggests that specific de-repression may be a common mechanism for controlling the expression of neuron-specific genes and that a common repressor protein may silence the expression of several such genes. Future efforts will be directed at cloning the gene for this repressor protein and understanding how it in turn is controlled.

Neural Development in Mammals and *Drosophila* Uses Similar Regulatory Molecules

We have used the chromaffin-neuron precursor cell lines to isolate new genes that may be candidates for controlling the development of these cells. Our approach is based on the idea that *Drosophila* genes controlling neural development might also be conserved in mammals. One important set of genes that control neuronal development in *Drosophila* are those of the *achaete-scute complex* (AS-C). These *scute* genes encode a group of related proteins that act by binding to DNA, thereby controlling the activity of other genes. We succeeded in isolating two *scute*-related genes from the rat chromaffin-neuron precursor cell line. The structures of these genes are remarkably similar to those of their fruit fly counterparts. Moreover, the rat *scute*-related genes appear to be expressed specifically in neuronal precursor cells, like their counterparts in the fly. These data indicate that there has been a striking parallel conservation of gene structure and cell type specificity during evolution. They further suggest that the *scute*-related genes may control the development of mammalian neurons. These exciting findings suggest that the molecular mechanisms controlling nerve cell development in vertebrate and invertebrate organisms may be fundamentally similar. We are now in the process of testing this hypothesis by making targeted mutations in these genes in mice.

Cell Fate Choices in the Nervous System and Elsewhere

Spyridon Artavanis-Tsakonas, Ph.D.—Investigator

Dr. Artavanis-Tsakonas is also Professor of Cell Biology and Biology at Yale University. He received his M.A. degree in chemistry from the Federal Institute of Technology (Eidgenoessische Technische Hochschule) in Zurich and his Ph.D. degree in molecular biology from the University of Cambridge, England, for work done at the MRC Laboratory of Molecular Biology. His postdoctoral work was done at the Biozentrum of the University of Basel with Walter Gehring and at Stanford University with David Hogness.

A fundamental issue in the biology of multicellular organisms is how a cell acquires its specific developmental fate. The molecular rules by which neighboring cells choose developmentally distinct paths are unknown. We are particularly interested in how these rules apply to the nervous system and thus have been exploring the molecular biology of early neurogenesis in the fruit fly.

The central nervous system in *Drosophila* derives from a set of precursor cells (neuroblasts) that segregate from the epidermal precursors (dermoblasts) in the very early ectoderm. Embryological, genetic, and molecular studies have demonstrated that neuroblast segregation depends on interactions between neighboring cells. Several genes capable of interfering with this process have been identified, and the *Notch* locus is central. As we have shown, *Notch* codes for a transmembrane protein with homology to the mammalian epidermal growth factor (EGF), implying involvement in cell-surface events.

For several years we have been studying *Notch* and other genetic elements in the cell interactions underlying neuroblast differentiation. Through these studies it became clear that the mechanisms governing the segregation of the neuroblasts from the epidermal precursors in the neural ectoderm are used not only in neurogenesis but also in a broad spectrum of tissues and cell fate choices.

In an attempt to examine the complexity of the genetic circuitry in which *Notch* is integrated and to identify genes whose products may interact directly with the *Notch* protein, we have been using genetic screens designed to identify suppressors or enhancers of specific *Notch* mutations. Through this analysis, five genes, which we operationally refer to as the "*Notch* group," have been identified, and extensive interactions among them documented. Besides *Notch*, the set includes four members that code for cell-surface, cytoplasmic, and nuclear elements. *Delta* and *Serrate* both code for transmembrane proteins with EGF homologous extracellular domains; *Enhancer of split* and *mastermind* code for nuclear proteins; and *deltex* appears to be cytoplasmic.

Among the various genetic interactions within the *Notch* group, those occurring between *Notch* and *Delta* most clearly suggest that they interact on the molecular level. To investigate this possibility, we examined the effects of *Notch* and *Delta* expression in *Drosophila* S2 cells. We found that *Notch*- and *Delta*-expressing cells formed mixed aggregates and that this occurs via their extracellular EGF homologous domains. Interestingly, *Notch*-expressing cells exhibit vesicular structures containing both *Delta* and *Notch* proteins, implying that *Delta* is internalized via *Notch* as a receptor. This relationship is currently being analyzed by electron microscopy.

Deletion mutagenesis of the extracellular domain of *Notch* determined that only two of the 36 EGF repeats are sufficient and necessary to mediate interactions with *Delta*. Thus the extracellular domain of *Notch* seems surprisingly modular and could potentially serve to bind a number of other proteins. Some of these may bind to subsets of the EGF repeats or to parts of the extracellular domain of *Notch*, while others may compete for the same binding sites. Indeed, *Serrate*, which displays homologies to the *Delta* protein, interacts with *Notch* in a *Delta*-like fashion and, since it binds to the same EGF repeats, may compete with *Delta* for binding to *Notch*.

On the basis of these results, it was proposed that *Notch* may function as a multifunctional receptor containing a series of ligand-binding sites, each of which may interact with more than one specific ligand. Such a model provides a plausible explanation for the pleiotropic action of *Notch* and suggests a general function for the gene throughout development.

If *Notch* acts as a receptor, how are *Notch*-mediated extracellular signals transmitted to the nucleus? This question is central to our current work and is being addressed at several levels.

Cloning and sequencing of *deltex* (which displays allele-specific interactions with *Notch*, *Delta*, and *mastermind*) revealed the primary structure of its product, a 700–amino acid protein with no known homologues. Preliminary results indicate that the *deltex* protein is cyto-

plasmic and appears to be influenced in its subcellular localization by the expression of *Notch*. While these observations are compatible with the notion that *Notch* interacts with a cytoplasmic protein, recent findings regarding both *Enhancer of split* and the intracellular domain of the *Notch* protein raise the possibility that *Notch* could directly interact with nuclear components.

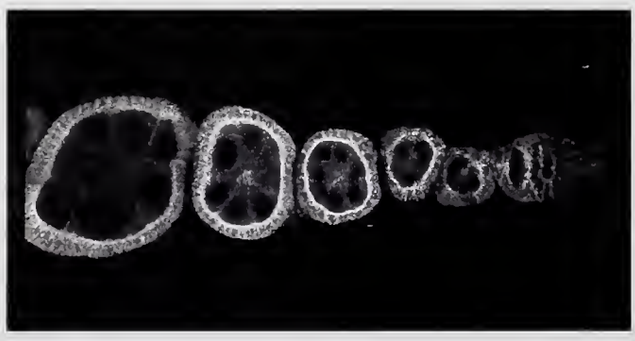
Most of the evidence supporting this hypothesis stems from our analysis of human *Notch* group homologues. In our attempts to dissect the molecular mechanisms involving *Notch* group members, we have been exploring their involvement in vertebrate development. While helping to establish the degree to which the *Drosophila* paradigm is valid for vertebrate species, this work also permits us to make use of the vertebrate model's experimental advantages, such as better-defined cultured cell lines.

The *Drosophila Enhancer of split* locus was shown to be defined genetically by several transcripts. With one exception, the genes contain the helix-loop-helix (HLH) motif found in a variety of transcriptional regulators, and they are partially redundant functionally. The single gene that does not contain the HLH motif displays homology to the β-subunit of the signal-transducing G proteins. Cloning of human homologues, which were named TLEs (for transducin-like *Enhancer of split*), revealed an extraordinary conservation between fly and human sequences. In addition, there seems to be a family of TLE genes in the human, though not the *Drosophila*, genome. Similarly, we found that the human genome contains at least two *Notch* homologous sequences.

Comparisons among the four isolated human TLE genes and the *Drosophila* counterpart showed that they harbor a structural motif implicated in nuclear-cytoplasmic transport. Referred to as the "CcN motif," it consists of a closely spaced combination of a nuclear localization sequence and potential phosphorylation sites for both casein kinase II and cdc2 kinase. Such a sequence is consistent with the notion that TLEs act as nuclear effector molecules. Indeed, immunocytochemical localizations show that both human and *Drosophila* molecules are predominantly nuclear but can also be detected in the cytoplasm.

Surprisingly, comparisons of the known *Notch* homologues, including the two human sequences, reveal the existence of a CcN domain in the intracellular part of *Notch*. We do not yet have any evidence regarding the functional significance of this motif, which raises the provocative possibility that *Notch* may possess some nuclear functions.

In conclusion, work in the past year extended our knowledge regarding the molecular rules underlying the action of the *Notch* group genes. Our studies—including analyses of expression patterns, genetic interactions, and mutant phenotypes—reinforce the notion that at least some of these genes code for elements of a pleiotropic cell interaction mechanism that is involved in cell fate choices in many different tissues, including the nervous system. Within those tissues, however, cell fate choices can also depend on tissue-specific interaction events. Our current working hypothesis is that regulative events controlling cell fate choices throughout development rely on the parallel action of tissue-specific mechanisms and on *Notch*-mediated interactions. *Notch* activity, while apparently general, can nevertheless be modulated by different ligands such as *Delta* or *Serrate*.



Expression of Notch protein in the early stages of wild-type Drosophila *egg chambers (from dissected ovaries) as shown by confocal microscopy. Notch is highly expressed at the apical surface of the follicle cells that surround the developing egg yolk.*

From Xu, T., Caron, L.A., Fehon, R.G., and Artavanis-Tsakonas, S. 1992. Development *115:913–922.*

The Complement System

John P. Atkinson, M.D.—Investigator

Dr. Atkinson is also Professor of Medicine and of Molecular Microbiology at Washington University School of Medicine and Physician at Barnes Hospital, St. Louis. He received his B.A. degree in zoology and M.D. degree from the University of Kansas. His training in internal medicine was at Massachusetts General Hospital, Boston, and at NIH. He conducted postdoctoral research at NIH and Washington University.

MANY poorly understood but relatively common and serious human diseases involve aberrations of the immune system. Such conditions range from mild forms of arthritis to life-threatening autoimmune illnesses such as systemic lupus erythematosus.

The immune system provides a powerful arsenal of proteins to attack and eliminate infecting microorganisms. A major means of accomplishing this task is through the special biologic partnership of antibodies and complement. Proteins from these two systems circulate in blood and are either formed in response to the organism (antibodies) or present at all times (complement).

The work in this laboratory centers on the complement system. Its name is derived from its function of "assisting" antibody in immune reactions. However, we now know that the complement system also serves as an independent immune system capable of attacking foreign elements by itself. This first-line defense appears to have occurred early in evolution, preceding antibody. The complement system consists of at least 30 proteins that interact with each other in reactions resembling a cascade or waterfall. The activated molecules that result from this scheme destroy the invading microbe and promote the inflammatory response. The complement system is a powerful, swift, and highly effective means to fight infectious organisms. As might be anticipated, an inherited deficiency of a complement component predisposes an individual to infections.

The production of antibodies is triggered when foreign substances invade the body. Although complement independently attacks the pathogen, the binding of antibodies to a target such as a bacterium initiates and potentiates a series of reactions in which complement proteins swarm onto the surface of the microbe. Such components serve as ligands for complement receptors on blood cells. A foreign particle so coated quickly becomes adherent to and is ingested by these blood cells. During complement activation, small fragments (peptides) that promote the inflammatory response are liberated from the complement proteins. These molecules dilate blood vessels and summon scavenger cells, called phagocytes, from the bloodstream. The phagocytic white blood cells, upon arrival at the site, find organisms that are already prepared for ingestion; i.e., they are coated with antibody and complement. This phenomenon is known as opsonization (from the Greek word *opsonein*, to prepare for the table). As a result, the infection is sequestered and eliminated via phagocytosis.

Sometimes, however, the immune system, instead of synthesizing antibodies to foreign materials, seemingly makes a mistake and produces antibodies that react with its own cells. For example, in certain immune disorders, individuals make antibodies to their own red blood cells. Antibody and complement then attach to the red blood cells and destroy them. Such conditions are known as autoimmune diseases, in this case autoimmune hemolytic anemia.

In other pathologic conditions, the immune system does not efficiently eliminate the infectious particles, and excessive quantities of immune complexes form. Immune complexes consist of antibodies, complement proteins, and the foreign particle. These immune complexes are proinflammatory and cause tissue damage if deposited in undesirable locations such as joints, skin, and kidney. This leads to arthritis, dermatitis, and glomerulonephritis, respectively.

Thus there are two general ways in which the humoral immune system can damage its own tissue. The first is to produce antibodies to self-components. The second is to form excessive amounts of immune complexes.

Our goal is to understand how immune complexes form and are processed. We are studying a biologic mechanism that evolved to remove immune complexes from the circulation. This process helps prevent the pathologic accumulation of these complexes in tissue. This system can be likened to an "inner space shuttle." The physiologic vehicle for this journey is the red blood cell. The red blood cell participates in this reaction through a complement receptor protein on its surface. These receptor proteins latch onto complement-coated foreign particles, such as vi-

ruses and bacteria. As the circulating red blood cell passes through the liver or spleen, its immune cargo is released and metabolized. The red blood cell then returns to the circulation ready to ferry another load.

Our laboratory is studying the complement receptor involved in this process. We are also examining the complement proteins that swarm onto the foreign particles. Furthermore, we are investigating certain "housekeeping" proteins of the complement system. Because of the powerful destructive capabilities of the complement system, it is perhaps no surprise that the body must keep this system tightly regulated. Special proteins are synthesized to protect the body's own cells from damage by complement factors. Our laboratory discovered a new family of genes that encode for at least six complement receptor and regulatory proteins. Two of these regulatory proteins occur on almost all cells of the body. These are termed decay-accelerating factor and membrane cofactor protein. Recently these regulatory proteins have been demonstrated to be expressed in relatively high concentration on reproductive tissue, including placental tissue and sperm. A new direction for the laboratory concerns the role of these proteins in reproduction. Modulating the function of these regulatory proteins may also be important in improving the destruction of tumor cells.

A further potential benefit from our research involves the field of transplantation. Presently the supply of donor organs is far less than the number of patients. Many people worldwide die while waiting for a suitable organ. Use of animal organs has not been possible, partly because of the complement system's attack on these transplants. However, this problem could be surmounted by using animal organs genetically engineered with complement regulatory proteins designed to prevent attack by the host's complement system. We are investigating these possibilities.

Finally, because infectious, autoimmune, and immune complex–mediated illnesses result from aberrations of the complement system, our research efforts are helping to define the pathophysiologic basis of such diseases. In many autoimmune and immune complex–mediated diseases, there is an inherited defect in the handling of immune complexes or in the activation or regulation of the complement system. Variations in the structure, function, and expression of complement proteins are important aspects of autoimmune diseases. The information gleaned from such research efforts could provide a basis for new treatments.

The Molecular Biology of Smell

Richard Axel, M.D.—Investigator

Dr. Axel is also Higgins Professor of Biochemistry and Molecular Biophysics and of Pathology at Columbia University College of Physicians and Surgeons. He received his undergraduate degree from Columbia College and his M.D. degree from the Johns Hopkins University School of Medicine. He then came to Columbia University as a resident in pathology at the College of Physicians and Surgeons. He held fellowships in Columbia's Institute of Cancer Research (in Sol Spiegelman's laboratory) and the Department of Pathology. Dr. Axel is a member of the National Academy of Sciences. Among his many honors are the Eli Lilly Award in Biological Chemistry and the Richard Lounsbery Award from the National Academy of Sciences.

PERIPHERAL neurons in vertebrate sensory systems respond to environmental stimuli and transmit these signals to higher sensory centers in the brain, where they are processed to allow the discrimination of complex sensory information. Mammals possess an olfactory sensory system of enormous discriminatory power. Humans, for example, are thought to be capable of distinguishing among thousands of distinct odors. Even subtle alterations in molecular structure of an odorant can lead to profound changes in perceived odor quality. How is this diversity and specificity accomplished? The detection of chemically distinct odors presumably results from the association of odorants with specific receptors on olfactory neurons that reside in specialized epithelium in the nose. The brain must distinguish which receptors or which neurons have been activated to allow the discrimination among different odorant stimuli. What mechanisms have vertebrates evolved to allow the recognition of this huge array of odorant molecules? How does the brain know what the nose is smelling? Insight into these problems is likely to depend on the isolation and characterization of the odorant receptors expressed in the nose.

We have recently identified an extremely large multigene family whose members are likely to encode a diverse family of odorant receptors. We have determined the sequence of 30 genes and have deduced the protein sequence of 30 different putative olfactory receptors. The olfactory proteins are clearly members of a superfamily of receptors that traverse the membrane seven times. Analysis of the proteins reveals structural features that may render this family particularly well suited for the detection of a diverse array of structurally distinct odorants.

These olfactory proteins can be divided into several different subfamilies that exhibit significant divergence in the region of the receptor molecule thought to interact with the odorous ligand. These observations suggest a model in which each of the individual subfamilies encodes receptors that bind distinct structural classes of odorants. Within a given subclass of odorous ligands, the members would recognize more subtle variations among odor molecules of a given structural class. As such, this superfamily of molecules would be uniquely suited to its putative role in the fine discrimination of odor molecules of subtly different structures.

The isolation of this large family of genes encoding the receptor molecules immediately provides one solution to the problem of olfactory perception. How do we recognize so diverse an array of odors? At one extreme, we would argue that the recognition of diverse odorants could be accomplished by a small number of promiscuous receptors, each capable of interacting with several structurally distinct odor molecules. Alternatively, olfactory perception would result from the presence of a large number of different receptor molecules, each capable of interacting with one or a small number of specific odorants. The size of the gene family we have characterized suggests that there are indeed a very large number of odorant receptors, each capable of interacting with a small number of odorous ligands. These observations are in sharp contrast to other sensory systems, such as vision or touch, where discrimination of sensory information is accomplished by a rather small number of receptor modalities.

How then does the brain distinguish which receptors or which neurons have been activated to allow the discrimination between different odorant stimuli? In other sensory systems, such as vision and touch, neurons in the brain are organized in a topographic map that identifies the position of a sensory stimulus. Thus the position of a given neuron in the brain is used to define the location of a sensory stimulus within the external environment. Olfactory processing does not extract spatial features of the odorant stimulus. Relieved of the necessity to encode information about the spatial localization of a sensory stimulus, the olfactory system may use spatial segregation of sensory input solely to encode the identity of the stimulus itself. Recent data utilizing the

receptor genes as molecular probes suggest a model of olfactory processing in which sensory neurons expressing distinct odorant receptors all project to a restricted region within the olfactory bulb, the first relay station in the brain. In this instance the discrimination of odors would be a consequence of the position of second-order neurons in the olfactory bulb, such that spatially localized groups of neurons would preferentially respond to different odorants. The molecular identification of the genes encoding a large family of olfactory receptors has provided initial insight into the logic of olfactory processing in the mammalian brain.

Mammalian Developmental Genetics

Gregory S. Barsh, M.D., Ph.D.—Assistant Investigator

Dr. Barsh is also Assistant Professor of Pediatrics at Stanford University. He received his M.D. and Ph.D. degrees from the University of Washington, where he studied inherited diseases of collagen biosynthesis in the laboratory of Peter Byers. Dr. Barsh's postgraduate training was in internal medicine and medical genetics at Harbor–University of California Hospital, Los Angeles, and at the University of California, San Francisco. His research in the laboratory of Charles Epstein focused on a molecular and genetic characterization of recessive lethal mutations at the mouse agouti *locus.*

VERY little is known about the genetic control of mammalian development. But embryogenesis of all mammals follows a similar plan, and the basic rules discovered in one species are likely to apply to others. By studying the mouse, a species in which the early embryo can be observed and manipulated, we will better understand how genes control human development and how disruption of these processes may lead to such abnormalities as miscarriages and birth defects.

In organisms traditionally subject to experimental genetic analysis, such as fruit flies and nematodes, mutations in a particular developmental pathway can be selected in a comprehensive screening experiment. In mice, however, this approach has been hampered by the inability to study and recover conditional mutations and by the inefficiency of generating new mutations through the insertion of mobile genetic elements. As a result, much of our insight into mammalian developmental genetics comes from the study of preexisting mutations. We are examining a group of previously identified genes that affect development around the time of implantation. In addition, we are developing a system to allow the conditional disruption of genes with recessive phenotypes in cell culture and in transgenic mice.

Characterization of the Mouse *agouti* and *kreisler* Genes

Located within a small region of mouse chromosome 2 are a group of genes required for fundamental aspects of peri-implantation development. The mouse *agouti* (*A*) locus, originally described as a coat color gene, controls the timing and type of pigment deposition in developing hair follicles. Several *A* mutations, including A^{y}, a^{16H}, and a^{x}, are lethal when homozygous, and at least three genes required for embryonic development have been defined at or near the *A* locus. Closely linked to *A* is the segmentation gene *kreisler* (*kr*), which affects formation of the inner ear by interfering with the number and pattern of metameric units, the so-called rhombomeres, in the developing hindbrain.

Toward the eventual goal of isolating these genes, we have constructed high-resolution genetic and physical maps of mouse chromosome 2, using classical and somatic cell genetic approaches. These maps are based on a variety of molecular markers in the area, including the insertion sites of two endogenous retroviruses, *Emv-15* and *Xmv-10*, which have been closely associated with the A^{y} and *a* mutations, respectively. Our results have shown that neither *Emv-15* nor *Xmv-10* has caused an *A* mutation. We have also established the order and relative distances between many closely linked molecular markers in the region and are beginning to lead toward candidate cDNAs affected by the *A* and *kr* mutations.

We have further characterized the relationships among rhombomere formation, segmentation, and the *kr* mutation, based on a detailed analysis of branchial arch derivatives in *kr/kr* adults. In collaboration with Michael Frohman and Gail Martin, we have studied the expression of rhombomere-specific genes in *kr/kr* embryos. Two of the rhombomeres affected by the *kr* mutation contribute to the second and third branchial arches, which ultimately form, among other structures, the hyoid bone. In *kr/kr* adults, the lower part of the hyoid, normally derived from the third branchial arch, exhibits some morphologic characteristics of the upper part of the bone, normally derived from the second branchial arch.

In *kr/kr* embryos, the expression domains of *Krox-20*, *Hox-2.6*, *Hox-2.7*, and *Hox-2.9*, which probably code for transcription factors, and *Int-2*, which codes for a growth factor and may play a role in morphogenesis of the inner ear, are all shifted rostrally toward the head. Taken together, these results suggest that *kr* affects the location of position-specific gene expression along the rostrocaudal axis and that the kr^{+} gene product plays a key role in the acquisition of positional identity. By isolating the *kr* gene and further characterizing the *kr/kr* phenotype, we hope to clarify

the molecular pathways of mammalian segmentation and the role of *Hox-2* genes in rhombomere morphogenesis.

Insertional Mutagenesis

An important development in mammalian experimental embryology has been the ability to isolate embryonic stem cells from preimplantation mouse embryos, which can be modified in cell culture and then used to reconstitute an intact animal. Like the whole organism, these cells are diploid and contain two copies of every autosomal gene. When embryonic stem cells are mixed with a fragment of "reporter" DNA, such as the coding regions of the neomycin resistance gene or the β-galactosidase gene, insertion of the exogenous DNA provides a "gene trap," in which expression of the reporter sequences is controlled by regulatory elements of an endogenous gene.

Insertion of the reporter DNA is likely to disrupt expression of the endogenous gene, but in most cases, expression from the uninterrupted homologue will be sufficient to prevent phenotypic effects. To block this expression in a single step, and in a conditional fashion, we have constructed a series of gene trap vectors that contain an inducible promoter on the strand opposite the reporter sequences. "Captured" cell clones are first selected by antibiotic resistance or β-galactosidase staining. After removal of selective pressure and activation of the inducible promoter, an antisense transcript is generated against coding sequences of the endogenous gene. In certain cases, this transcript will function in trans, thereby inhibiting expression of the endogenous gene from the uninterrupted homologue.

Control experiments suggest that these DNA vectors are capable of trapping endogenous genes and that basal expression of the inducible promoter does not reduce appreciably the frequency of trapping. We are currently isolating a panel of captured cell clones. The efficiency of these antisense promoters in cis will be analyzed by measuring expression of the reporter sequences before and after antisense induction. Potential effects of trans inhibition will then be tested by examining the phenotypes of chimeric mice that contain the mutant embryonic stem cells. This approach will allow the phenotypic effects of a recessive mutation to be studied in a diploid organism by altering only one of the two copies and may be applicable to many organisms and developmental systems.

Ultimately we plan to place the gene trap vectors in the context of retroviral packaging elements to allow their introduction into the mouse germline by infection of developing gametes. Theoretical advantages of using retrotransposons as insertional mutagens include an enzymatic mechanism of integration that conserves host DNA sequences, accessibility to most if not all segments of the genome, and the efficient detection and recovery of new integration sites. These advantages, however, have been difficult to achieve in the mouse, in part because of the low efficiency of retrotransposition in the germline.

In the case of endogenous ecotropic proviruses and exogenous Moloney murine leukemia virus (MoMuLV), a block to retroviral infection of developing gametes seems likely to be due to a combination of factors, including decreased accessibility of germline cells to viral particles with very short infectious half-lives and a transcriptional block to expression from the retroviral long terminal repeat (LTR). As a first step toward overcoming these hurdles, we have constructed, in collaboration with Patrick Brown (HHMI, Stanford University), transgenic mice that contain the transcriptional control sequences for the mouse protamine 1 gene fused to the coding sequences for MoMuLV. Because the mouse protamine 1 gene is highly expressed in postmeiotic spermatids, the transgene should provide a rich source of native retroviral particles in the vicinity of developing male gametes. Our preliminary results show that the chimeric protamine-MoMuLV transgene is functional and can lead to retroviral infection of somatic and germline tissues.

David H. Beach, Ph.D.—Investigator

Dr. Beach is also Senior Staff Scientist at Cold Spring Harbor Laboratory and Adjunct Associate Professor in the Department of Microbiology at the State University of New York at Stony Brook. He received his undergraduate degree at the University of Cambridge, followed by a Ph.D. degree at the University of Miami, where he worked with Marcus Jacobson. His postdoctoral studies were done with Sydney Shall at the University of Susex and with Amar Klar at Cold Spring Harbor Laboratory.

THE cell division cycle interests an increasing number of scientists and continues to be the central focus of our research. The two most critical events of the cycle are the replication of DNA that occurs during the S (synthetic) phase and the segregation of chromosomes into daughter cells that occurs in the M (mitotic) phase. DNA replication and mitosis are profoundly different cellular events, both molecularly and in terms of cellular mechanics, but both are regulated by closely related enzymes, members of a family known as the cyclin kinases.

Each of these kinases has a catalytic subunit, of which cdc2 is the prototype, and a regulatory subunit that is a cyclin. Cyclins derive their name simply from the property by which they were first recognized. Their abundance oscillates in the division cycle, with the same periodicity as the cycle itself. Many cyclins have been discovered, all related in their amino acid sequence. In humans there are at least five groups of such proteins, called cyclins A, B, C, D, and E. Within each group there may be further members, such as D1, D2, D3 or B1, B2. Cyclins of different classes appear to act at different stages of the cell cycle—B-type cyclins at mitosis, A type at S phase, and so on.

Also, in the case of the D-type cyclins, each member of the family displays a specific tissue distribution. D-type cyclins have attracted particular interest because they appear to act early in the cell division pathway and can also mutate to give rise to oncogenic forms. With cyclin D1, the gene is amplified or rearranged in a wide range of human tumors, including cancers of the breast, esophagus, bladder, and lymphatic system. It is not yet known how overexpression or activation of cyclin D1 contributes to the pathology of cancer, but a clear connection has been established.

A major activity in the Beach laboratory has centered on identification of the regulatory molecules controlling cyclin kinases, particularly the cyclin B/cdc2 enzyme that functions in mitosis. This enzyme is influenced by at least three phosphorylation events, some activating and others inhibiting. Tyrosine phosphorylation of cdc2 is the best-characterized inhibitory mechanism, and the relevant tyrosine kinases and phosphatases have been identified both in yeast and human cells.

The tyrosine kinases are encoded by the *mik1*$^+$ and *wee1*$^+$ genes of yeast, and the tyrosine phosphatases by the *cdc25*$^+$ gene. Kinases and phosphatases act antagonistically. Thus, in the absence of the tyrosine kinase, cells enter mitosis prematurely before the completion of DNA replication. Lack of the tyrosine phosphatase causes arrest in the G_2 phase of the division cycle.

A further class of negative regulators of mitosis has recently been discovered. The *pim1*$^+$ (premature initiation of mitosis) gene of yeast encodes a homologue of a human protein called RCC1. In the absence of *pim1*/RCC1 function, cells also enter mitosis before the completion of DNA replication. The gene product *pim1*/RCC1 is an integral nucleosomal protein (present at one copy per nucleosome) and is responsible for signaling the state of the chromatin to the cdc2/cyclin B enzyme. Only when the chromatin is fully replicated can the cell proceed to division by activation of the cyclin B/cdc2 enzyme. The mechanism by which the *pim1*/RCC1 nucleosomal protein communicates with the cyclin kinases is an area of major current interest.

Many substrates of the cyclin kinases have been discovered. They include major structural proteins such as lamins and vimentin, which play a vital role in the structural reorganization of the cell that occurs during mitosis. Other substrates include transcription factors and also negative growth regulators such as the p53 and rb proteins.

It is unclear why such different cellular activities as DNA replication and cell division are regulated by a family of cyclin kinases that are clearly related evolutionarily. The evolutionary similarity implies that there was once only one cyclin kinase and thus that this enzyme regulated only one cell cycle event in a primitive cell. Was that event DNA replication or cell division? It is very difficult to imagine the primitive cell, because all

known cells are exquisitely interdependent entities. For example, the replication of DNA is dependent on proteins, but the proteins are encoded by DNA, and so on. It is clear, however, that the primitive cyclin kinase has evolved in present-day cells into a very complicated family of enzymes that regulate all stages of the cell division cycle.

Genetic Control of Morphogenesis in Drosophila

Philip A. Beachy, Ph.D.—Assistant Investigator

Dr. Beachy is also Assistant Professor in the Department of Molecular Biology and Genetics at the Johns Hopkins University School of Medicine. He did his graduate work in the Department of Biochemistry at Stanford University School of Medicine with David Hogness. Before joining the Hughes Institute at Johns Hopkins, he spent two years as Staff Associate at the Carnegie Institution's Department of Embryology. He has been the recipient of several fellowships, including a Sloan Foundation Fellowship in Neuroscience.

TWO conceptually distinct processes operate during *Drosophila* embryogenesis. In the first of these, a detailed system of spatial information is generated by a cascade of interactions among several groups of genes and their products. During morphogenesis, the second phase, interpretation of the blueprint results in the complex arrangement of structures seen in the larva and adult.

The first process is initiated during oogenesis by maternal input of broad spatial cues. The interacting genes that receive and refine this information in the embryo have been extensively characterized, in most cases at the molecular level. Much less is known about how morphogenesis is directed, though it is generally believed to occur through regulation of target genes by the localized proteins generated during construction of the blueprint. This laboratory's current interests center upon two groups of genes that operate at the interface between these processes, the segment polarity genes and the homeotic genes.

The *hedgehog* Segment Polarity Gene

The segment polarity group is the latest acting of the segmentation gene groups. It functions in the refinement of positional information at the level of individual cells within embryonic segments. During the earlier, syncytial period of *Drosophila* embryogenesis, the absence of cell membranes permits diffusion between nuclei, which can account for many of the interactions between early acting segmentation genes and their products. Some segment polarity genes, however, are expressed and function long beyond the syncytial stages. Communication between cells for refinement of positional information requires signaling across cell membranes. Consistent with a role in signaling, the products of many segment polarity genes are secreted or localized to nonnuclear compartments.

During the past year our laboratory isolated and characterized the segment polarity gene *hedgehog*. The *hedgehog* locus was identified in the exhaustive genetic screens carried out by Christiane Nüsslein-Volhard, Eric Wieschaus, and their colleagues, and was named for the bristle pattern observed in mutant larvae. The phenotype resembles that of a number of other segment polarity genes, including *wingless*.

The *wingless* gene is expressed within each segment in a stripe of cells adjacent to that in which *engrailed*, another segment polarity gene, is expressed. The normal function of each gene is required within its own stripe for maintenance of the other gene's expression within the neighboring (but not overlapping) stripe. The mutual requirement of each gene for the other's continued expression in a neighboring cell implies a mechanism for communication between these cells. Since *wingless* encodes a secreted peptide factor and *engrailed* a homeodomain-containing transcription factor, additional components must be involved in the signaling mechanism.

Interest in this pathway has been stimulated by recent results elsewhere indicating that targeted inactivation of the murine homologues of *wingless* and *engrailed* produce specific developmental defects of the midbrain and cerebellum. Other genes that may be involved in this pathway are *hedgehog* and *patched*, which from genetic evidence appear to interact. While *patched* encodes a protein with apparent multiple membrane-spanning domains, the product of the *hedgehog* gene had not previously been characterized. We gained entry into the *hedgehog* locus with the aid of an "enhancer trap" P-transposable element inserted near the gene. Analysis revealed a sequence capable of encoding a protein of 471 amino acid residues. This includes a region near the amino terminus suggestive of targeting to the secretory pathway, either as a secreted polypeptide or as an integral membrane protein.

Either possibility would permit interaction of the *hedgehog* protein with the membrane-spanning *patched* protein, in accordance with the genetic prediction. Also consistent with such an interaction, the *hedgehog* gene is expressed coincidently with *engrailed,* while *patched* is expressed coincidently with *wingless* in adjacent cells. The physical nature of the interaction re-

mains to be characterized, along with the intracellular consequences of this signaling event.

Regulatory Functions of Homeotic Genes

Homeotic genes play a primary role in conversion of the spatial information generated by the earlier-acting segmentation genes into morphological structure. The products of homeotic genes direct morphogenesis in a manner that produces the diversity of structures distinguishing the segments. This task is presumably accomplished by differential regulation of target genes, though the identities of these targets and the mechanisms by which homeotic genes achieve their biological specificity remain largely unknown.

We have studied in detail the DNA sequence-recognition properties of homeodomains from the homeotic genes *Deformed* and *Ultrabithorax*. While these domains are very closely related, we have demonstrated clear differences in DNA sequence preferences that lead to moderate differences in affinity for individual binding sites. Our recent work demonstrates that these differences in sequence preference map to the two DNA-contacting portions of the homeodomain. Even small differences in binding to individual sites can lead to large differences in regulatory effect when multiple sites are present. This is due to cooperative binding of homeodomain proteins to multiple sites, a property we have studied extensively for the *Ultrabithorax* protein. Differences in DNA sequence preference by homeotic gene products may thus account for a substantial part of the biological specificity of homeotic gene function.

Molecular Studies of Human Genetic Disease

Arthur L. Beaudet, M.D.—Investigator

Dr. Beaudet is also Professor in the Institute for Molecular Genetics and the Departments of Pediatrics and Cell Biology at Baylor College of Medicine. He received his B.S. degree in biology from Holy Cross College and his M.D. degree from Yale University. After completing his pediatric residency at the Johns Hopkins Hospital, he performed postdoctoral research at NIH.

EXTRAORDINARY progress in the understanding and diagnosis of human genetic diseases has occurred over the past decade. Much of this progress was made possible by the application of recombinant DNA techniques to the analysis of those diseases caused by an alteration in a single gene. There are thousands of such disorders, including well-known conditions such as cystic fibrosis, Duchenne muscular dystrophy, hemophilia A and B, sickle cell anemia, and Tay Sachs disease. Although this progress in the diagnosis of single-gene disorders continues unabated, the ability to treat genetic disease virtually stands still by comparison. There is hope that at least some of these diseases will yield to somatic gene therapy.

The extraordinary diagnostic ability that allows anticipation of genetic symptoms in individuals and their offspring raises important societal opportunities and concerns. These include prevention of genetic disease through reproductive planning, treatment prior to development of symptoms, and possible discrimination (as for insurance or employment) on the basis of one's genetic predispositions.

Abnormalities of a single gene, however, are not responsible for most of the common medical problems of adult life. Rather, these are caused by differences in a number of genes in combination with environment and life-style. Problems of this type include atherosclerosis, hypertension, diabetes mellitus, and autoimmune disease. Work in our laboratory involves various aspects of precise genetic diagnosis, efforts to develop somatic gene therapy, and attempts to understand the complex genetic basis of some of these multifactorial disorders.

Cystic Fibrosis

Cystic fibrosis (CF) is a common genetic disease affecting approximately 1 in 2,500 Caucasians. Most CF patients die of progressive lung disease during childhood or young adult life. About 1 in 25 Caucasians carry an abnormal CF gene, and about 1 in 625 couples are at high risk of having an affected child. Our laboratory has been actively involved in developing and implementing DNA testing for purposes of prenatal diagnosis and carrier detection of CF. Prenatal diagnosis for high-risk couples who have a 1-in-4 chance that a pregnancy will be affected with CF is now a routine matter, and such diagnosis is performed at many laboratories throughout the world. Similarly, carrier detection for close relatives of CF patients is routine.

Much more controversial is whether carrier testing for CF should be offered to most or all couples prior to reproduction. Although a single common defect is present in the majority of CF chromosomes, the remaining fraction of abnormal chromosomes contains dozens of different mutations. One goal of efforts in our laboratory is to develop efficient methods for detection of such multiple abnormalities. Currently carrier testing identifies 80–90 percent of carriers and would detect up to 80 percent of couples at risk for having a CF child. Although it is unclear whether CF carrier testing should be offered to all couples immediately or in the very near future, it is our view that carrier testing will be the starting point for multiphasic DNA testing to provide information regarding reproductive and health risks.

Precise genetic diagnosis for CF, and for other genetic disorders, does not guarantee an effective treatment. It would be desirable to have an animal model for easier analysis and therapeutic trials. As one step in the process of developing an animal model in mice, we have characterized the sequence of the mouse CF gene. We have prepared DNA clones that are suitable for disrupting the normal mouse gene (which is similar to the human gene) in cultured embryonic stem (ES) cells. These altered ES cells can be used to generate mice affected with CF. The work to obtain a CF mutant mouse is supported in part by a grant from the Cystic Fibrosis Foundation.

Somatic gene therapy for CF focuses on the lung, since it is involvement of the lung that proves fatal for the vast majority of CF patients. Treatment of the lung presents a unique circumstance, for therapeutic agents can be delivered by

inhalation rather than injection. We are preparing to evaluate various methods for pulmonary gene therapy. Liposomes containing DNA, retroviral vectors, and adenoviral vectors will be used for direct instillation into the lungs or for various aerosol deliveries, beginning with normal rodents and perhaps progressing to primates.

Leukocyte and Endothelial Cell Adhesion Molecules

Leukocytes (white blood cells) are involved in defense against infection and in most inflammatory responses. Through complex mechanisms, the various forms of leukocytes roll along the blood vessel walls and eventually migrate through the endothelial layer. This ability is mediated to a large extent by carefully regulated adhesion molecules on the surface of leukocytes and endothelial cells.

We are attempting to understand the role of these proteins in normal biology and in multifactorial genetic diseases. This involves preparing mutant mice and analyzing naturally occurring genetic variation in humans. It is quite possible that variation in these genes could be an important factor in the predisposition to multifactorial disorders such as atherosclerosis, autoimmune disease, and a wide range of inflammatory processes, including arthritis and diabetes mellitus.

We are focusing initially on CD18, which is a subunit of leukocyte integrin, and on intracellular adhesion molecule-1 (ICAM-1), a member of the immunoglobulin gene family. CD18 is found on the surface on many leukocytes, and its adhesiveness can be sharply increased through an activation process. ICAM-1 is found on many cells, particularly endothelial cells, where it can bind the CD18-containing integrins and mediate firm cell attachment and migration of leukocytes out of the bloodstream.

Individual genes can be altered in mice by the technique of homologous recombination (exchange of DNA between related molecules). Using this methodology in ES cells, a mutation was introduced into the mouse CD18 gene, and the alteration was transmitted through the germline so that the altered animals could be bred. The mutant animals have a partial rather than complete impairment of the CD18 gene's function. Leukocyte function shows significant alteration in these animals, and they will be studied for any differences in inflammatory responses or susceptibility to atherosclerosis.

A mutation has also been introduced into the ICAM-1 gene, and mice carrying this mutation in the majority of their cells are being bred to obtain transmission of the defect to the germline. Efforts have also been initiated to prepare a mutation in the P-selectin gene, whose product is expressed on platelets and endothelial cells. Adhesion proteins of the selectin family bind to sugar molecules, rather than to proteins, on other cells. The work to obtain mouse mutations in leukocyte and endothelial cell adhesion molecules is supported in part by a grant from the National Institutes of Health.

A defect in CD18 causes a rare human genetic disease. In its severe form, this condition, known as leukocyte adhesion deficiency (LAD), leads to fatal susceptibility to infections. LAD is an attractive model for development of somatic gene therapy for bone marrow–derived cells. A recombinant retrovirus encoding the human CD18 protein has been prepared and has been used to transfer the gene into mouse and human cells. Bone marrow transplantation was performed in mice, and the infected cells express human CD18 when reimplanted into the mouse. Cultured cells from LAD patients were infected with retrovirus, and the adhesion properties of the cells were restored to normal. Bone marrow cells from affected patients are being cultured in the laboratory and infected with the retrovirus in preparation for performing human gene therapy.

Genetic variation in the adhesion molecules of human leukocytes and endothelial cells is being identified. In the ICAM-1 gene, for example, there is an amino acid difference that is prevalent in the population. It is of interest to determine whether different forms of the ICAM-1 gene in the human population are associated with different susceptibilities to inflammatory diseases and atherosclerosis. Some of the long-term goals of this project are to identify naturally occurring genetic variations that alter susceptibility to common disease processes. In addition, the mutant mice can be used to determine if decreased function of some of these cell adhesion molecules reduces chronic inflammatory disease processes (e.g., autoimmune disease, arthritis, multiple sclerosis, diabetes mellitus). If reduced gene function were to lessen the disease processes in mouse models, this would provide evidence that the function of these proteins might be blocked by drugs as a means of slowing the analogous processes in humans.

Molecular Genetics of Diabetes Mellitus

Graeme I. Bell, Ph.D.—Investigator

Dr. Bell is also Professor of Biochemistry and Molecular Biology, Medicine, and Genetics at the University of Chicago. He received his B.S. degree in zoology and M.S. degree in biology from the University of Calgary, Canada. He earned his Ph.D. degree in biochemistry and biophysics from the University of California, San Francisco, where he also did his postdoctoral research. Before moving to the University of Chicago, Dr. Bell served as Senior Scientist at the Chiron Corporation. Dr. Bell has received a number of awards for his work, including the Outstanding Scientific Achievement Award from the American Diabetes Association, the Rolf Luft Award from the Swedish Medical Society, and the Distinguished Alumni Award from the University of Calgary.

DIABETES mellitus is a disorder of carbohydrate metabolism characterized by elevated blood glucose levels. In the United States, an estimated 6 million persons are known to have diabetes, and there are probably an equal number with unrecognized disease. About 10 percent of persons over age 65 have diabetes, and the complications of the cardiovascular, kidney, visual, and nervous systems are major causes of morbidity and mortality.

Clinically diabetes is a heterogeneous disorder. One form, insulin-dependent diabetes mellitus, primarily affects children and adolescents. It results from immunological destruction of the insulin-producing cells of the pancreas; because of the absolute deficiency of insulin, patients require insulin therapy for survival. The more common form of diabetes, non-insulin-dependent diabetes mellitus (NIDDM), includes about 90 percent of diabetic patients. This form results from reduced insulin levels in some patients and a relative deficiency in others and is due to abnormal functioning of the insulin-producing cells or decreased responsiveness of tissues to insulin. In these patients the blood glucose levels can usually be controlled by diet or by drugs that can be taken orally to improve insulin secretion and action.

As with other common diseases of middle age such as cardiovascular disease and hypertension, genetic factors contribute to the development of NIDDM. Our goal is to identify the genes that increase the risk of developing diabetes and to determine how, together with environmental and life-style factors, they result in the elevated blood glucose levels that define this disorder.

We have taken an approach that combines genetics and molecular biology. Our working hypothesis is that a relatively small number of potentially identifiable major genes increase the risk of developing diabetes and that the individual's overall genetic background, together with environmental and life-style factors, influences the phenotypic expression of the major susceptibility genes.

In our genetic studies, we are studying diabetes-prone families in which NIDDM has an early age-at-onset and a clear autosomal dominant mode of inheritance. In one such family we have identified a DNA marker on the long arm of human chromosome 20 that is tightly linked to the diabetes-susceptibility gene. We are presently trying to isolate this gene and to identify the mutation that impairs its normal function and leads to failure of the insulin-secreting pancreatic β-cells. The identification of this gene may suggest others whose mutation could also cause NIDDM.

Drawing on our understanding of the pathophysiology of NIDDM, we are also cloning and characterizing genes that might reasonably contribute to diabetes susceptibility. Since the pancreatic β-cell plays an important role in the pathogenesis of all forms of diabetes mellitus, we are particularly interested in studying the genes that are responsible for its unique physiological characteristics. These include genes encoding hormones such as insulin and islet amyloid polypeptide (a newly discovered hormone-like peptide of uncertain function), ion channels, and receptors for hormones and other agents that regulate insulin secretion. In addition to cloning and characterizing genes expressed in β-cells, we are also identifying DNA polymorphisms in these genes to facilitate genetic studies of their contribution to development of NIDDM. Furthermore, we are studying the effects of hyperglycemia and other altered metabolic states on their expression in order to determine if changes in the levels of the proteins encoded by these genes can provide a molecular explanation for the failure of the β-cell to secrete appropriate amounts of insulin in response to glucose in patients with NIDDM.

Recent studies of the receptor for the polypeptide hormone somatostatin are beginning to shed new light into the mechanism by which this hormone inhibits insulin secretion. In addition, they are leading us into the area of neurobiology and hormonal regulation of neuronal function. Somatostatin is a cyclic 14–amino acid polypeptide

found throughout the gastrointestinal tract, including the pancreas, and nervous systems. It has many diverse functions, including inhibiting the secretion of insulin and glucagon from the cells of the pancreatic islets and of growth hormone from the pituitary. In the central nervous system, somatostatin acts as a neurotransmitter to regulate neuronal activity and to facilitate the release of other neurotransmitters. It may also play a role in centrally mediated behaviors such as movement and cognition. We have recently cloned the genes encoding three somatostatin receptors. All are members of the seven-transmembrane-spanning, GTP-binding protein family of receptors. One of these, SSTR2, is expressed in the pancreatic β-cell and is likely to be responsible for mediating the inhibition of insulin release by somatostatin. The three somatostatin receptor isoforms that we have characterized have unique pharmacological properties and couple to different intracellular effector systems, thus providing at least a partial explanation for the diverse physiological effects of somatostatin. We are continuing our studies of this family of receptors to gain a better understanding of how somatostatin binding results in inhibition of insulin secretion.

In addition to genes expressed in the pancreatic β-cell, we have also characterized genes for membrane proteins involved in the transport of glucose across the plasma membrane. Our studies have revealed unexpected functional complexity that could have important implications for the treatment of diabetes. These studies indicate that facilitative glucose transport is not the property of a single protein but rather involves a family of at least five structurally related proteins. These proteins have distinct patterns of tissue distribution and different kinetic properties and are independently regulated. These features allow the precise disposal of dietary glucose under varying physiological conditions. Recent studies have shown that one of the five proteins we have identified is the major glucose transporter of neuronal cells and that the major function of another is to transport fructose across the plasma membrane. This fructose transporter is expressed on the luminal surface of the absorptive epithelial cells of the small intestine and kidney tubules and is also present in early and late spermatids in the testes. Sperm require a fructose transporter, because they utilize fructose in seminal fluid as an energy source.

Our studies are leading to a better understanding of the causes of diabetes mellitus. In addition, they are providing new insight into the function of somatostatin in the regulation of neuronal function. The identification of the fructose transporter in sperm may also have implications in reproductive physiology. Our results also illustrate how studies in one area can impact those in another, because of the exquisite manner in which the organism is able to utilize similar proteins for different physiological functions.

Development of the Drosophila *Peripheral Nervous System*

Hugo J. Bellen, D.V.M., Ph.D.—Assistant Investigator

Dr. Bellen is also Assistant Professor in the Institute for Molecular Genetics, the Division of Neuroscience, the Department of Cell Biology, and the Program in Developmental Biology at Baylor College of Medicine. Educated in Belgium, he received a degree in commercial engineering and began research in sociometry, but decided to pursue a career in medical science. He obtained a D.V.M. degree from the University of Ghent and a Ph.D. degree in genetics from the University of California, Davis, where, in John Kiger's laboratory, he studied mutations that affect behavior in the fruit fly. Later, in Basel, Switzerland, as a postdoctoral fellow with Walter Gehring, he helped to develop the enhancer detection system.

OUR research is centered on the development of the peripheral nervous system (PNS) of the fruit fly *Drosophila melanogaster*. The results will help us analyze the development of the nervous system of many eukaryotic species. We have focused on the PNS of *Drosophila* because its cells are relatively easy to study and because sophisticated genetics can be applied to this model organism.

We are presently studying the *couch potato* (*cpo*) and *neuromusculin* (*nrm*) genes, which are expressed at the onset of embryonic nervous system development and later in embryogenesis in most cells of the PNS. These genes were identified in enhancer detector screens in which regulatory sequences of genes are identified by means of a β-galactosidase reporter. Here we describe the preliminary characterization of these essential genes and the development of a novel technique to ablate nervous system cells.

The *cpo* gene was so named because homozygous mutant flies are viable but have poor jump responses, poor flight abilities, slow recovery after ether anesthesia, and are generally hypoactive. More than 10 insertional *cpo* mutations were recovered from several enhancer detector screens; none of the insertions caused a complete loss of function of the gene. Some insertional alleles cause recessive embryonic lethality, but no defects were observed in the PNS of the mutant embryos.

A molecular analysis of *cpo* was initiated with the isolation of *cpo* cDNAs. *In situ* hybridizations to whole-mount embryos showed that the gene is expressed in the precursor cells of the PNS as well as glia of the PNS and CNS (central nervous system). Sequencing data show that the gene encodes at least three different proteins that contain many stretches of the same amino acids. Two of these proteins contain a domain that may allow binding to RNA molecules. This binding domain is found in many proteins that are involved in processing immature RNA in a form that can be readily translated into a protein. The CPO protein has most significant homology with two other RNA-binding proteins that are found in nervous tissue: HuD, a human brain paraneoplastic encephalomyelitis antigen, and elav, the protein from a *Drosophila* nervous system–specific gene (*embryonic lethal abnormal vision*).

The *cpo* gene sequence was used to produce a polyclonal antibody to CPO. The antibody allows the localization of this nuclear protein in the developing embryo, in essentially the same cells as the transcript. In addition, the antibody recognizes a protein that is associated with polytene chromosomal bands isolated from the salivary gland of third instar larvae. This association may be mediated through the binding of CPO to nascent RNA that is being synthesized from the DNA of the giant polytene chromosomes. Thus it is possible that CPO acts as an essential PNS differentiation factor by controlling some aspects of RNA maturation specific to the cells in which CPO is expressed.

In addition to its biological functions, the *cpo* gene is also a hot spot for P-element enhancer detector insertions. Ten insertions that cause a variety of phenotypes were mapped at the nucleotide level, and molecular defects in *cpo* mutant strains that carry an enhancer detector were determined. All insertions recovered so far are clustered in a 200-bp genomic fragment that contains key regulatory regions of the *cpo* gene. Most insertions are integrated upstream of the first nucleotide of the longest cDNA but downstream of the consensus binding sites of three known *Drosophila* transcription factors. Thus they are probably located between the enhancer-binding sites and the promoter of *cpo*. Five insertions causing different phenotypes have inserted at exactly the same location. There is a perfect correlation between the observed defects (or their lack) in embryos or adults and the size and the orientation of these insertions—e.g., large insertions are less detrimental than smaller insertions. These data suggest that the distance between the binding sites of the transcription factors and the transcription initiation site is critical for proper *cpo* regulation. In addition, insertions with the *lacZ* gene

in the opposite orientation of *cpo* transcription cause much less severe defects than insertions in the same orientation. These observations provide a unique opportunity to study the interactions between enhancers and promoters *in vivo*. The study of *cpo* is also supported by the Muscular Dystrophy Association.

The *nrm* gene is expressed early in the development of the PNS and muscles. It maps at cytological band 80A and was also identified in an enhancer detector screen. To our knowledge only a single insertion has been identified in *nrm*, and the β-galactosidase expression pattern in embryos of this strain is virtually identical to that of *cpo* during early development. Flies homozygous for this particular insertion are homozygous viable. A mutagenesis experiment allowed us to recover 10 recessive lethal mutations that fail to complement each other. Although many subtle defects have been observed in different mutant strains, no consistent morphological defects have been observed.

The cloned *nrm* gene encodes a single 4.6-kb transcript that is expressed transiently in the sensory mother cells and developing neurons and support cells of the PNS (5–10 h). Later in embryonic development (14–17 h) the transcript is also observed in most embryonic muscles. The earliest expression of *nrm* in muscles coincides with the formation of the neuromuscular junction. Sequencing data show that the *nrm* gene encodes a novel protein that contains a signal peptide and eight immunoglobulin domains. The protein is possibly anchored in the membrane of the cells by a secondary modification of the protein. On the basis of these molecular data and the expression pattern of *nrm*, we propose that *nrm* is a novel neural cell adhesion molecule that may play an important role in growth cone guidance in the PNS and may affect the formation of the neuromuscular junction. These hypotheses are presently being tested.

To study the role of glial cells in the development of the PNS, we are developing a method that allows temporally induced arrest of most functions in specific cells. These arrests are induced by temperature shifts of flies that express temperature-sensitive (ts) forms of diphtheria toxin (DT-A^{ts}) under the control of cell- or tissue-specific regulatory sequences. DT-A^{ts} genes were isolated in a mutagenesis screen using the yeast *Saccharomyces cerevisiae* and subsequently tested in PNS neurons, namely the R1–R6 photoreceptor cells of transgenic fruit flies. Four DT-A^{ts} have been partially characterized in yeast cells, and three have been tested in *Drosophila*. These toxins show similar temperature dependence, suggesting that they may be useful in a wide range of species. We are presently trying to ablate glial cells in the PNS and CNS, in an attempt to examine the role of glia in the living organism.

Expression pattern of the couch potato protein in an almost fully developed Drosophila *embryo. This protein is expressed in all cells of the peripheral nervous system and in a subset of cells of the central nervous system. Anterior, left; dorsal, up.*

Research and photograph by Sandra Kooyer and Diana D'Evelyn in the laboratory of Hugo Bellen.

Genetic Manipulation of Hematopoietic Stem Cells

John W. Belmont, M.D., Ph.D.—Assistant Investigator

Dr. Belmont is also Assistant Professor of Molecular Genetics, Pediatrics, and Microbiology and Immunology at Baylor College of Medicine. He received his undergraduate degree from the University of Texas, Austin, and his M.D. and Ph.D. degrees from Baylor College of Medicine, where he worked with Robert Rich. After internship and residency training in pediatrics at Children's Hospital, Washington, D.C., he completed a fellowship in medical genetics at Baylor College of Medicine.

PLURIPOTENT hematopoietic stem cells are the "seed" for the development of all blood cells. These cells, which normally reside in the bone marrow, arise in early fetal development and persist throughout adult life. They can be removed from the bone marrow and transplanted into a prepared recipient; they will then stably reconstitute the entire blood and immune systems. We have developed techniques for the efficient transfer of genes into mouse and human stem cells, using retroviral vectors. The long-term goal of these studies is to perfect methods that could be used for the treatment of various genetic and acquired diseases.

A viral vector system based on the Moloney murine leukemia virus (MoMuLV) has been chosen because of its potential for very high gene transfer efficiency. The unique life cycle of this retrovirus makes it attractive for adaptation as a vector, since the foreign genetic material is stably integrated into the host cell genome. MoMuLV vector particles are able to carry their genetic material to the target cells but are incapable of replicating and spreading as a live infectious agent.

We have used two model systems to investigate properties of the stem cells. In one model the bacterial antibiotic resistance gene *neo* introduces distinct genetic tags into individual stem cells. The second model uses the human enzyme adenosine deaminase (ADA) as the molecular marker for gene transfer. This system is particularly suitable for studies of expression of genes by the retroviral vectors. In addition, the genetic deficiency of ADA causes a form of severe combined immune deficiency, so that successful laboratory experiments with this gene may in time facilitate the clinical application of the gene transfer procedures.

Our earlier work demonstrated that genes could be introduced into hematopoietic stem cells but that the process was much less efficient than in the more mature cells of the marrow. In mouse transplant experiments, only about 50 percent of the animals retained expression of the human ADA enzyme in their blood for more than six months. This has led to an investigation of the conditions in cell culture that would optimally support the survival or proliferation of the stem cells.

In collaboration with Doug Williams (Immunex, Seattle), we have evaluated the effects of several recombinant hematopoietic growth factors on retroviral vector–mediated gene transfer into stem cells. These factors included interleukin-3, -6, and -7; granulocyte colony-stimulating factor (G-CSF); and leukemia inhibitory factor (LIF). LIF has been of special interest because among its many biological functions, it appears to prevent the differentiation of mouse embryonic stem cells. If it had a similar action on hematopoietic stem cells, it might allow the preservation of their developmental capacity in culture.

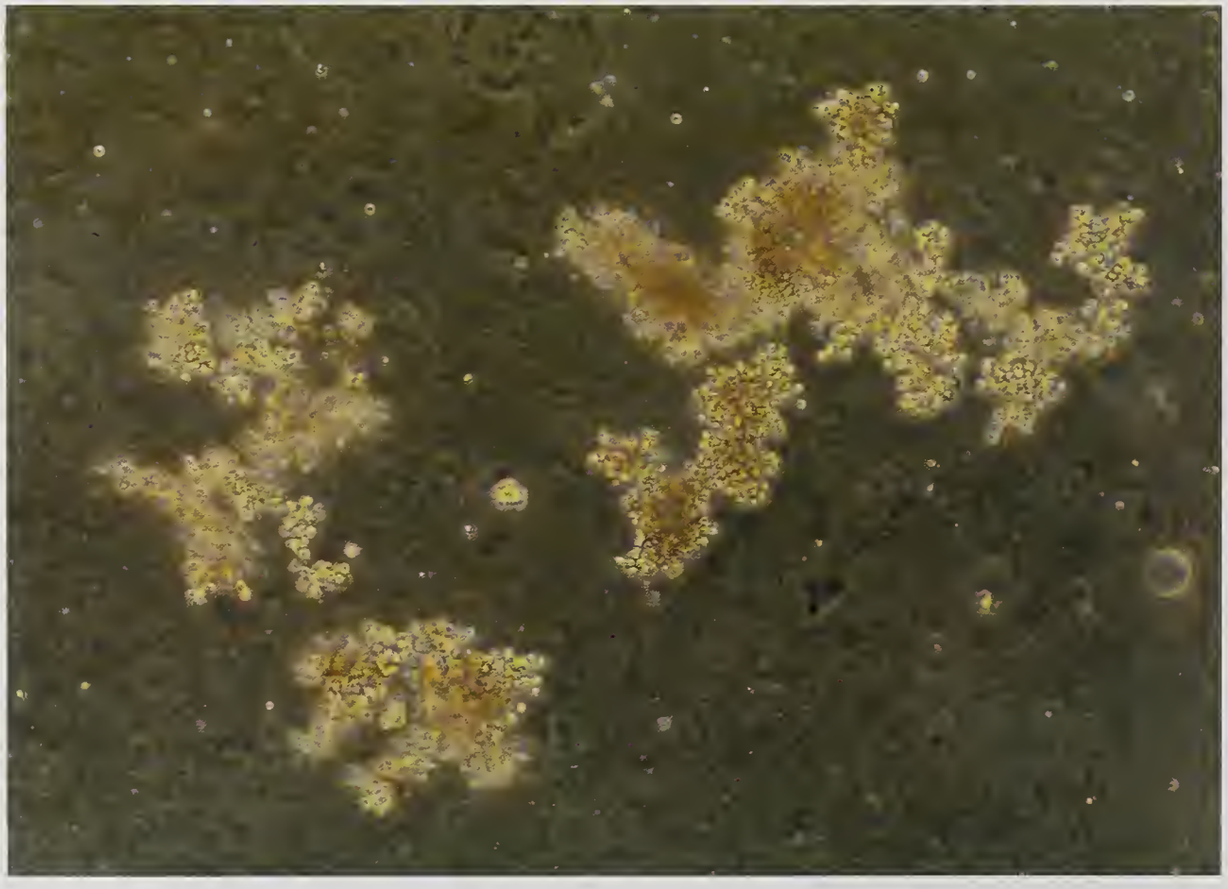
A novel assay using inbred transgenic mice was used to test the activity of LIF on stem cells. These experiments indicate that LIF preserves the stem cells during the culture period required for gene transfer. Inclusion of LIF in the bone marrow cultures allows about 70 percent of the stem cells to be infected with the retrovirus. Subsequently all the mice transplanted with such cells maintain high-level expression of human ADA in all their blood and immune system organs for at least six months. This finding has recently been supported in collaborative experiments with Savio Woo (HHMI, Baylor College of Medicine) using a retroviral vector that encodes the human α_1-antitrypsin gene.

We suspect that LIF acts in concert with one or more other growth factors in our experimental model. The recently isolated ligand for the c-*kit* cell surface receptor is a promising candidate for a factor that causes stem cells to replicate. In an effort to survey critical growth factor/ligand systems expressed by stem cells, we have tried a cloning method based on the conservation of gene sequence to identify new candidate receptors. Messenger RNA extracted from stem cell–enriched fractions of bone marrow has been used as a template for amplification of conserved receptor tyrosine kinase sequences. Among the clones are three that had not previously been described. We are currently sequencing these genes

to determine if they are novel receptors involved in the regulation of hematopoiesis.

To analyze better the behavior of stem cells and their progeny in culture and after transplant, we are testing a new method for introducing unique genetic identifiers into individual stem cells. This method allows identification of the stem cells by the polymerase chain reaction (PCR), which is especially suited for analysis of very small numbers of cells. A family of vectors carrying the bacterial gene *neo* have been constructed that are identical except for small variations in size. This size variation is conveniently distinguished by PCR. When mixed together these vectors provide an array of potential markers that can integrate into each stem cell. Because each stem cell receives one to four vectors, the integration of a particular subset of the vectors provides a genetic fingerprint unique to each stem cell family. Preliminary results with these vectors indicate that we can mark the most primitive hematopoietic precursors very efficiently. We plan to use this method to compare different culture conditions for stem cells.

Retroviral vector genetic marking may also prove to be very informative in clinical bone marrow transplantation. Collaborative studies with Albert Deisseroth (M.D. Anderson Hospital, Houston) investigated whether the techniques that have been used to optimize mouse bone marrow stem cell gene transfer could be used for human cells. These experiments demonstrate the importance of supporting cells in the culture (called stromal cells) for efficient gene transfer. The work was performed using an assay of human stem cells that depends on their persistence in culture. We hope that many valuable lessons will be learned from these studies that will contribute to the long-term goal of treating disease by the use of therapeutic vectors.



Human bone marrow colony called an erythroid burst or BFU-E. The colony grows from a single cell to form 3–9 clusters of red blood cells. Measurement of gene transfer into such colonies has helped to improve gene therapy protocols.

Research and photograph by Kateri Moore in the laboratory of John Belmont.

Proteins of the Spectrin-based Membrane Skeleton

G. Vann Bennett, M.D., Ph.D.—Investigator

Dr. Bennett is also Professor of Biochemistry at Duke University Medical Center. He received his M.D. and Ph.D. degrees from the Johns Hopkins University Medical School. He completed postdoctoral training at Harvard University in membrane protein biochemistry with Daniel Branton. Before joining Duke University, he was on the faculty in the Department of Cell Biology and Anatomy at Johns Hopkins.

STRUCTURAL proteins in the cytoplasm and membranes of cells provide the basis for spatial organization of the diverse components of eukaryotic cells. These proteins thus are principal participants in fundamental cellular activities such as cell motility and cell-cell interactions. Our work over the past 10 years has focused on plasma membranes. We initiated these studies in the human erythrocyte. This relatively simple cell has provided an experimentally accessible model system for detailed dissection of protein-protein interactions that are responsible for the structure and organization of the plasma membrane.

The principal structural protein in the erythrocyte membrane is the flexible rod-shaped molecule spectrin, which is organized in a two-dimensional network attached to the cytoplasmic surface of the plasma membrane. Spectrin molecules are attached at their ends to form a series of hexagons and pentagons that closely resembles a geodesic dome. The binding of spectrin to the protein ankyrin attaches the spectrin network to the plasma membrane. Ankyrin also interacts with high affinity with the cytoplasmic domain of an integral membrane protein (a protein that traverses the membrane and actually has portions exposed on both the inner and outer membrane surfaces). The spectrin-based membrane network or skeleton is required for normal stability of erythrocytes in the circulation. Abnormalities in amounts or function of spectrin and associated proteins result in hemolytic anemias and are the basis for diseases such as hereditary spherocytosis and hereditary elliptocytosis.

Proteins closely related to spectrin are present in many vertebrate cells and are associated in most cases with the plasma membrane. Spectrin is present in especially high amounts in brain, where it comprises 3 percent of the total membrane protein. The spectrin-based membrane skeleton in brain and other tissues is likely to play an important role in providing organization of integral membrane proteins in the plasma membrane and for coupling membrane proteins to elements of the cytoskeleton. Potential physiological consequences of these activities include stabilization of the lipid bilayer and organization of membrane proteins in specialized regions on the cell surface in polarized cells.

Specific aims of this laboratory are to elucidate the proteins in erythrocytes and other cells that mediate interaction of spectrin with membranes, determine how these protein interactions are regulated, and understand the cellular functions of the spectrin skeleton.

Ankyrins in the Nervous System

Ankyrin appears to function as an adapter between certain membrane proteins and the spectrin skeleton. We have discovered that brain contains multiple forms of ankyrin, with diversity due to distinct genes as well as alternative splicing of RNAs encoded by the same gene. We have determined the complete amino acid sequence of the major form of ankyrin in human brain and have discovered an unusual alternative form of this protein that contains a large inserted sequence. The large form of brain ankyrin is the first ankyrin detected during brain development and is targeted to neuronal processes including unmyelinated axons.

The same gene that encodes ankyrin in erythrocytes also is expressed in brain. This form of ankyrin is localized in the plasma membranes of certain neurons and is abundant in cerebellum, brain stem, and spinal cord. The erythrocyte form of ankyrin is missing in a strain of mutant mice developed at the Jackson Laboratory. Ankyrin-deficient mice experience degeneration of Purkinje cells, a major type of neuron in the cerebellum, and develop a stagger and difficulty in walking. Neurological problems of ankyrin-deficient mice may have counterparts in humans with slowly progressive diseases due to death of nerve cells.

Another isoform of ankyrin is highly concentrated along with the voltage-dependent sodium channel at the nodes of Ranvier of nerve axons. Nodes of Ranvier are specialized regions on the axons of nerves where the myelin or insulation of the axon is interrupted and where ions can enter

or leave the axon through ion channels. Localization of the voltage-dependent sodium channel at nodes of Ranvier is important for normal conduction of nerve impulses. We hope to identify the gene encoding the form of ankyrin at the node of Ranvier and eventually to understand the role of nodal ankyrin in organization of this specialized membrane domain. These studies will have relevance to diseases of neurons such as multiple sclerosis, where the myelin coating of axons is lost and sodium channels are no longer restricted to the nodes of Ranvier.

Diversity of ankyrins suggests that this family of proteins may interact with many membrane proteins. Characterized membrane proteins in brain that associate with ankyrin in *in vitro* assays include the voltage-dependent sodium channel and sodium/potassium ATPase. We have used the membrane-binding domain of the major form of brain ankyrin to isolate ankyrin-binding proteins and have identified a family of ankyrin-binding proteins found in plasma membranes of neurons and glial cells. These ankyrin-binding proteins represent 0.3 percent of adult brain membrane protein and appear late in postnatal development. An important goal for future work will be to isolate the cDNAs encoding these proteins and determine their function in adult brain.

Ankyrin-Independent Membrane Attachment Sites for Spectrin

Brain spectrin can also associate directly with membrane proteins through an interaction that is independent of ankyrin. We have discovered that calcium, in concert with three different calcium-regulated proteins (calmodulin, a calcium-activated protease, and protein kinase C), inhibits the direct spectrin-membrane linkage but has no effect on spectrin-ankyrin interactions. These results suggest that the spectrin skeleton includes both stable, ankyrin-mediated linkages and dynamic calcium-sensitive associations that are subject to metabolic control. Identification of the spectrin "receptor" is the first step in understanding the role of this type of spectrin-membrane interaction in cells.

Ankyrin Structure

Ankyrin contains three independently folded domains: one that interacts with certain membrane proteins, another that associates with spectrin, and a third that regulates associations of the binding domains. Surprisingly, the membrane-binding domain of ankyrin includes an amino acid sequence that is homologous to regions of sequence in a group of apparently unrelated proteins from flies, yeast, and even viruses. We have recently discovered that this conserved portion of the ankyrin sequence is responsible for the interaction of ankyrin with at least one membrane protein. Ankyrin thus contains a highly conserved and ancient structural motif that may have a general role in molecular recognition. We hope to determine the three-dimensional structure of this portion of ankyrin, with the expectation that this structure will help us understand interactions of ankyrin and other proteins with related sequences.

Adducin

The protein adducin is a candidate to play a role in assembly of the spectrin skeleton in erythrocytes, brain, and certain epithelial tissues. We have found that adducin is localized at sites of cell-cell contact in epithelial tissues. Adducin and spectrin are colocalized at cell contact sites and may be arranged in a structure analogous to the spectrin network of erythrocytes. The association of adducin with cell-cell contact sites occurs before assembly of other types of specialized cell junctions such as desmosomes.

Our working hypothesis is that adducin promotes assembly of a stable spectrin network at sites of cell-cell contact. A further hypothesis is that the spectrin network is an essential precondition for assembly of specialized cell junctions. Formation of appropriate cell-cell contacts and cell junctions is an essential event in embryogenesis and is one of the processes that is disturbed in diseases such as cancer. We are excited by the possibility that adducin and spectrin may participate in such a fundamental activity of cells. We have determined the complete protein sequence of both adducin subunits and are in the process of determining a physical model for organization of domains and subunits in the adducin molecule. Future experiments will evaluate the role of adducin and spectrin in formation of junctions between cells.

TNF and the Molecular Pathogenesis of Shock

Bruce A. Beutler, M.D.—Associate Investigator

Dr. Beutler is also Associate Professor of Internal Medicine at the University of Texas Southwestern Medical Center at Dallas. After receiving his M.D. degree from the University of Chicago (Pritzker School of Medicine), he served as an intern and resident at the Southwestern Medical Center. His postdoctoral fellowship with Anthony Cerami was completed at the Rockefeller University, which he left as an assistant professor to assume a position at the Southwestern Medical Center.

THIS laboratory studies basic mechanisms that lead to septic shock, a serious condition arising as a result of many types of infection. We have learned that the final common pathway leading to shock involves the production of certain cytokines, particularly tumor necrosis factor (TNF), by host cells known as macrophages and by other cell types as well.

Once TNF has been produced, it alters the metabolism of cells throughout the body, triggering a breakdown of protein and fat stores. If TNF is chronically secreted at low levels, a state of wasting called cachexia will develop. This condition is seen in cancer and many other forms of chronic illness. On the other hand, if massive quantities of TNF are released over a short time, as in widespread injury, the protein activates neutrophils and endothelial cells in such a way that shock occurs.

Because TNF is a critically important molecule in various human disease processes, we have sought to understand how its biosynthesis is controlled. Probably the most potent stimulus for TNF release is a molecule known as lipopolysaccharide, or endotoxin. This molecule is produced by gram-negative bacteria, which have a remarkable tendency to cause shock. In the course of a gram-negative infection, endotoxin is released into the bloodstream. It is harmless to most cells but is a powerful activator of monocytes and macrophages, triggering their release of TNF with all of its attendant consequences.

By studying different portions of the TNF gene, we have shown that endotoxin causes two separate responses within the macrophage: 1) It causes increased transcription of the TNF gene, leading to a marked accumulation of TNF mRNA within macrophages. 2) It causes far more efficient translation of the mRNA—i.e., increases the speed with which the mRNA is read to produce TNF protein. Acting in concert, these two effects are responsible for a 10,000-fold increase in the rate of TNF biosynthesis and thus a massive net effect.

For a number of technical reasons, it has been very difficult to demonstrate the major sources of TNF in living animals. It is not clear, for example, whether it is made by normal tissues in healthy animals or whether such "baseline" production is important for maintenance of physiological or immunological processes. Similarly, it is not clear whether the TNF that arises in cancer is derived from cells of the tumor or from host cells that act in response to the tumor. TNF is believed to be made in a variety of autoimmune or allergic diseases, but again, the principal source of the protein remains uncertain.

To address these questions, our laboratory has produced transgenic mice that express a reporter construct in which an easily measurable enzyme (chloramphenicol acetyltransferase, or CAT) is employed as a marker for TNF. In other words, CAT synthesis also occurs in cells that produce TNF. CAT remains confined, however, to the cell of origin, whereas TNF is secreted and becomes widely dispersed in the organism. Using these animals, we have found that during normal development TNF is made by cells of the thymus. Other investigators have further reported that thymic production of TNF is essential for normal development. Although the protein does not appear to be produced elsewhere in healthy animals, it is readily induced by administration of lipopolysaccharide or by various authentic infections.

A second tissue in which CAT synthesis marks TNF production is the placenta, the organ that is formed in pregnant mammals to nourish the developing fetus. The placenta contains cells of two separate origins: part of the placenta, the decidua, is derived from the mother. The remainder of the placenta, the trophoblast, is derived from the fertilized egg. Studies utilizing the reporter construct have effectively proved that the TNF gene is constitutively expressed by the trophoblast. TNF may also be constitutively expressed by trophoblastic tumors. Its function within the placenta remains to be established; however, it is interesting to consider that the protein may play an important role in normal pregnancy.

Our laboratory has also made progress in understanding the mechanism of action of drugs that inhibit TNF biosynthesis and in devising molecules that block the action of TNF once it has

been released. These studies might lead to better therapies for shock and other disorders. Glucocorticoid hormones (e.g., prednisone, dexamethasone, and cortisol) have long been used as anti-inflammatory drugs. One of their principal effects appears to be a blockade of TNF biosynthesis, which depends upon inhibition of both TNF gene transcription and mRNA translation. Other drugs of a class known as phosphodiesterase inhibitors (e.g., theophylline, caffeine, and pentoxifylline) also block TNF biosynthesis, achieving their effect by preventing TNF mRNA accumulation. They appear to function at a different locus than do glucocorticoids. We have shown that the two classes of drugs when combined exert a synergistic effect.

Recently the cell-surface receptor for TNF was cloned in a number of laboratories. We have engineered a recombinant molecule in which the TNF receptor is attached to a portion of a normal antibody, yielding a new protein molecule in which two TNF-binding sites are expressed. This bivalent TNF-binding protein strongly inhibits the biological effects of TNF, is highly stable *in vivo*, and may be produced in large quantities by recombinant techniques. We anticipate that this type of molecule will allow a thorough investigation of the many effects of TNF in health and disease and may also be useful as a therapeutic tool. Of particular interest will be its use in studies of TNF production and action in the thymus and in the placenta.

Cytotoxic T Lymphocyte Recognition

Michael J. Bevan, Ph.D.—Investigator

Dr. Bevan is also Professor of Immunology at the University of Washington, Seattle. He received his Ph.D. degree for work performed at the National Institute for Medical Research, Mill Hill, London. He did his postdoctoral work at the Salk Institute in the laboratory of Melvin Cohn, after which he was a faculty member in the Center for Cancer Research and the Department of Biology at the Massachusetts Institute of Technology. He later conducted research in immunology at the Scripps Research Institute before moving to the University of Washington. Dr. Bevan was elected Fellow of the Royal Society of London.

TWO functionally different types of T lymphocytes mature in the thymus and populate the peripheral lymphoid organs. Helper T lymphocytes, the first type, respond to antigen by releasing lymphokines, which activate macrophages or augment the response of antibody-producing B cells. Cytotoxic T lymphocytes, on the other hand, when induced with antigen, specifically lyse target cells expressing the antigen and release interferon-γ. Their function is thought to be crucial in the response to intracellular pathogens, such as viruses and bacteria, and they may eliminate some tumor cells.

T lymphocytes can only recognize antigen presented in association with cell surface glycoproteins encoded by the major histocompatibility complex (MHC). Recent work has shown that cytotoxic T cells recognize short (usually nonameric) antigenic peptides presented in the groove of class I MHC molecules.

Since class I molecules are expressed on virtually all tissues, any cell type can be a target for cytotoxic cells. The peptides presented by these MHC molecules derive from degraded intracellular proteins, so all, or most, of a cell's normal components can provide class I–binding peptides. Normally, however, the T cells tolerate these self-peptide/self-MHC complexes. Viral infection, on the other hand, leads to the production of new proteins in the cell. These may provide nonamer peptides that can combine with class I and be presented on the surface. If cytotoxic T lymphocyte surveillance works well, these foreign-peptide/self-MHC complexes will be recognized and the virus-infected cell destroyed.

The class I MHC proteins expressed on the cell surface have three components: class I heavy chain, β_2-microglobulin, and a tightly bound peptide. This trimolecular complex is assembled shortly after synthesis of the heavy and light chains in a pre-Golgi compartment.

Many of the peptides that are bound to class I actually derive from cytosolic and nuclear proteins. It seems likely that cytosolic proteins are degraded in the cytosol by complex structures called proteosomes and that the peptide degradation products pass into the endoplasmic reticulum (ER). Some of the components of proteosomes are encoded within the MHC in close proximity to a pair of genes that belong in the superfamily of ATP-dependent transport proteins. These gene products may mediate the translocation of peptide antigen from the cytoplasm into the ER lumen.

Antigen Presentation-Defective Variants

A number of mouse and human cell lines have been described that synthesize normal class I heavy and β_2-microglobulin chains but neither assemble these chains for surface expression nor present endogenous antigens to cytotoxic T cells. We provided evidence that the defect in the mouse cell line RMA-S maps to chromosome 17, the chromosome encoding the MHC. When this cell line was fused with a wild-type partner, expression of MHC antigens was restored. When the wild-type chromosome 17 was selected against, expression of the MHC antigens was lost. John Monaco (Virginia Commonwealth University, Richmond) had previously identified and cloned *Ham-1* and *Ham-2*, two ABC transporter genes from the mouse MHC. In collaboration with his group and with James Forman and Kirsten Fischer Lindahl (HHMI, University of Texas Southwestern Medical Center at Dallas), we were able to show that transfection of the *Ham-2* gene into RMA-S restored surface expression of class I MHC antigens, as well as the ability to present five endogenous antigens on the surface for recognition by cytotoxic T lymphocytes. Thus a defect in the *Ham-2* gene leads to loss of surface class I expression. In conjunction with previous experiments with the human defective cell lines, this suggests that both *Ham-1* and *Ham-2* may be essential for class I surface expression, and in fact they may operate as heterodimers.

In other experiments we have shown that the RMA-S cell line expresses a subset of endogenous peptides on the surface with class I. For example, the octameric peptide derived from the vesicular stomatitis virus nucleoprotein is presented to cy-

totoxic T lymphocytes by RMA-S. After viral infection or transfection of the nucleoprotein gene alone, this epitope appears on the cell surface. By moving this sequence of eight residues around, we hope to be able to discover what makes these escape peptides special—is it the peptides themselves or surrounding sequences? The leakiness in this cell line also suggests that a *Ham-1* homodimer may function inefficiently in this mutant cell line.

Cytotoxic Responses to Intracellular Bacteria

Listeria monocytogenes is a frequent foodborne pathogen that causes severe disease in immunocompromised individuals. It is a gram-positive facultative anaerobe. Following uptake by macrophages, *Listeria* escapes from the hostile environment of the phagolysosome by secreting a hemolysin, listeriolysin, which ruptures the phagosome membrane. Once into the cytoplasm, the bacterium replicates and can push itself into a neighboring cell by polymerizing actin. Previous workers had demonstrated a cytotoxic T cell response in mice to *Listeria* infection. Our main priority was to determine which of the approximately 4,000 proteins made by *Listeria* were providing class I–associated peptide epitopes for cytotoxic T lymphocytes.

We determined that the secreted listeriolysin molecule itself provided a strong class I–restricted epitope in the mice we studied. To pinpoint the presumed nonamer within the 600 amino acid residues of the listeriolysin molecule, we used class I peptide-binding motifs that had predicted tyrosine at position 2 and leucine at position 9. The listeriolysin contained three nonamers that fit this criteria. These and others were made as synthetic peptides, one of which targeted our cytotoxic T lymphocytes at picomolar concentration. Thus we have used these motifs for the first time and shown that they are valid in pinpointing cytotoxic T lymphocyte epitopes.

In more recent studies we have been able to show that cytotoxic T lymphocyte immunity to this short 9-residue peptide presented by a class I molecule is sufficient to confer adoptive immunity to the whole bacterium. Lines of cytotoxic T lymphocytes with specificity only to this epitope, when transferred into syngeneic mice, confer protection against lethal doses of *Listeria monocytogenes*.

The listeriolysin epitope of *Listeria* that induces conventional cytotoxic T cell immunity is restricted by a classical class I molecule. "Classical" refers here to the highly polymorphic H-2K, D, or L loci in mice and their equivalent HLA-A, B, and C loci in humans. We have evidence that other peptides derived from this bacterium are presented by nonclassical, nonpolymorphic class I molecules. We know that class I is involved in presenting these peptides, because cell lines that lack β_2-microglobulin are unable to present the peptides. Furthermore, we know that the class I molecule maps just outside the MHC, as revealed by experiments with congenic mice differing only in this region.

But unlike the classical class I molecules, this restricting element is rather nonpolymorphic: i.e., cells from most mouse strains are able to bind and present the peptide. We have yet to identify the peptide derived from *Listeria* or the class I molecule involved. It will be interesting in the future to determine whether cytotoxic T lymphocytes directed to these peptides, seen in the context of nonpolymorphic MHC molecules, confer protective immunity.

Vascular Endothelium in Inflammation and Metastasis

Michael P. Bevilacqua, M.D., Ph.D.—Associate Investigator

Dr. Bevilacqua is also Associate Professor of Pathology and a member of the division of Cellular and Molecular Medicine at the University of California, San Diego. He received his undergraduate degree in biology from the University of Pennsylvania and his M.D. and Ph.D. degrees in pathology from the State University of New York at Brooklyn. As a postdoctoral research fellow with Michael Gimbrone, Jr., and Ramzi Cotran at Brigham and Women's Hospital and Harvard Medical School, he initiated work on endothelial leukocyte adhesion molecules. He continued these studies as a faculty member at Harvard Medical School before assuming his present position. Dr. Bevilacqua is a Pew Scholar in the Program in the Biomedical Sciences.

VASCULAR endothelial cells form a cobblestone-like lining of blood vessels throughout the body. Despite accurate predictions by certain investigators in the 18th century, the highly responsive and changeable nature of vascular endothelium was not recognized until the early 1980s. Endothelium was widely regarded as a "nonstick" surface that prevented blood-clotting and resisted the adhesion of circulating leukocytes. It is now apparent that the simple structural features of vascular endothelium belie a complex functional nature.

Our research has focused on the mechanisms by which the vascular endothelium can regulate inflammation, immunity, thrombosis, and tumor metastasis. It was demonstrated that soluble protein mediators known as cytokines could act directly on endothelial cells to increase the expression of prothrombotic activities and promote the adhesion of blood leukocytes. Investigation into the mechanisms of leukocyte-endothelial adhesion led to the identification and characterization of two endothelial cell-surface glycoproteins that were designated endothelial leukocyte adhesion molecule-1 (ELAM-1) and inducible cell adhesion molecule-110 (INCAM-110). This discussion focuses on endothelial adhesion molecules and their potential roles in human disease processes.

Endothelial Adhesion Molecules in Inflammation

The study of inflammation has one of the richest histories in biomedical research. Particularly noteworthy is the work of Julius Conheim (1839–1884), which provided one of the first (and best) microscopic descriptions of the process. He carefully observed blood vessels after a local injury in transparent membranes such as the tongue of a frog and described the vasodilation, edema, and leukocyte emigration with great accuracy. He also suggested that the blood vessel lining in the area of inflammation became sticky for leukocytes. This concept was largely lost during the next century, in which a theory of leukocyte response to soluble mediators (chemotactic factors) dominated.

Until the early 1980s endothelial cells were thought to move politely aside or be destroyed as leukocytes exited blood vessels to fight bacterial infections or contain foreign substances. By 1985 the view of vascular endothelium had begun to change. Recombinant cytokines were found to act directly on endothelial cells to increase dramatically the adhesion of blood leukocytes, and it became apparent that the endothelium produces a "sticky" surface. By 1987 the nature of that surface was being elucidated, with the identification of two cytokine-inducible endothelial adhesion molecules designated ELAM-1 and ICAM-1 (intracellular adhesion molecule-1). Studies on human tissues demonstrated that these molecules are expressed on endothelium at sites of inflammation.

A solid working hypothesis emerged: foreign substances such as bacterial products introduced into tissues would stimulate local cells such as macrophages to synthesize and secrete cytokines. The cytokines would then act on vascular endothelium to promote the expression of adhesion molecules that would bind passing leukocytes. Leukocytes could then exit the blood vessel and migrate toward the foreign substances following chemotactic gradients. By 1990, through the efforts of multiple laboratories, five endothelial cell-surface molecules that participate in leukocyte adhesion and inflammation had been identified, cloned, and characterized. These molecules fall into two families based on their structures.

A newly described family, the vascular selectins, contains three related molecules, two of which can be found on endothelial cells: E-selectin (a new name for ELAM-1) and P-selectin (previously known as GMP-140 or PADGEM). The second group of endothelial leukocyte adhesion molecules are part of a much larger family known as the immunoglobulin superfamily.

The pattern of expression of endothelial adhesion molecules as well as the specificity of their binding interactions with leukocytes contributes

importantly to the control of inflammatory processes. Acute inflammation is of relatively short duration, lasting as little as a few minutes or as long as 1–2 days. It is characterized by the exudation of plasma proteins and the emigration of leukocytes, especially neutrophils. Chronic inflammation is of longer duration and is associated with the extravasation of blood lymphocytes and macrophages and with the morphological changes in blood vessels and connective tissues.

In most cases the pattern of E-selectin expression correlates well with acute inflammation. This observation is consistent with the demonstration that E-selectin supports the adhesion of essentially all neutrophils, but only a small portion of lymphocytes. In contrast, INCAM-110 (known also as VCAM-1) supports the adhesion of lymphocytes but not neutrophils, and its pattern of expression is consistent with its primary role in chronic inflammatory processes.

The inflammatory process is essential for host defense. In certain settings, however, it can also contribute to debilitating and life-threatening diseases. Examples of human diseases with significant inflammatory/immunological components are adult respiratory distress syndrome (ARDS), myocardial reperfusion injury, rheumatoid arthritis, and various autoimmune diseases. Understanding the mechanisms of endothelial-leukocyte adhesion may allow us to enhance our defense mechanisms, or dampen their responses when appropriate.

Vascular Selectins and Their Carbohydrate Ligands

Each of the three selectins is a cell-surface glycoprotein with a mosaic structure. The portion of the molecule that is external to the cell, and hence reaches out into the bloodstream, contains a lectin domain. Lectins are protein molecules that bind to carbohydrates (sugars). Our laboratory is utilizing the recombinant E- and P-selectin proteins and soluble synthetic carbohydrates to determine the structures involved in recognition.

Early studies had suggested that E-selectin mediates the adhesion of blood leukocytes through the binding of a carbohydrate known as sialyl-Lewis X. This structure contains a terminal sialic acid as well as a fucose group bound to a carbohydrate backbone of galactose and *N*-acetylglucosamine. Surprisingly, we demonstrated that a soluble form of sialyl-Lewis X was only a modest inhibitor of E-selectin–dependent adhesion, whereas a related carbohydrate called sialyl-Lewis A, differing only in the linkage of the backbone sugars, was much more effective. Sialyl-Lewis A is not found on blood neutrophils but is expressed by a variety of cancers, including those of the colon. The relative blocking activities of sialyl-Lewis X and sialyl-Lewis A were reversed for P-selectin.

In addition, a third carbohydrate compound designated CD65, which is found on leukocytes and some cancer cells, was shown to be an effective blocker of adhesion to both E- and P-selectin. Subsequent studies using modified carbohydrates have demonstrated that the presence and precise arrangement of sialic acid and fucose are important in the recognition by the two selectins. Furthermore, minor modifications in the carbohydrate structures can result in greatly increased activity in blocking leukocyte adhesion.

A Possible Role in Cancer Metastasis

If cancers grew and invaded tissues locally but failed to metastasize, most of them could be cured by surgery. Indeed, the lethal effects of a cancer are generally ascribed to its metastatic abilities. Hematogenous metastasis involves release of cells from a primary tumor into blood vessels and transport to new tissues. Substantial evidence indicates, however, that the blood is hostile to cancer cells and generally destroys them. Thus most cancer cells must exit the bloodstream in order to survive. Our laboratory has focused on the mechanisms by which tumor cells bind the vessel wall and extravasate. We now appreciate that many tumor cells can interact with the vascular endothelium through the same molecules blood leukocytes use.

As noted above, colon cancers can express carbohydrate ligands for E-selectin. In a parallel line of investigation, we have demonstrated that human melanomas can express the integrin $\alpha 4\beta 1$, which is a receptor for the endothelial cell-surface molecule INCAM-110 or VCAM-1. These observations suggest the possibility of novel therapeutic approaches to block the metastatic spread of cancer.

In summary, vascular endothelium can exist in different functional states. These states are determined by stimuli such as cytokines and bacterial products. Activated endothelial cells express adhesion molecules for circulating leukocytes and can thereby regulate inflammatory processes. In addition, it appears that cancer cells can utilize these same adhesion molecules during the process of hematogenous metastasis.

Structural Studies of Molecules Involved in the Immune Recognition of Infected Cells

Pamela J. Bjorkman, Ph.D.—Assistant Investigator

Dr. Bjorkman is also Assistant Professor in the Division of Biology at the California Institute of Technology and Adjunct Professor of Biochemistry at the University of Southern California School of Medicine, Los Angeles. She received a B.A. degree in chemistry from the University of Oregon and a Ph.D. degree in biochemistry and molecular biology from Harvard University, where her thesis advisor was Don Wiley. She completed a low-resolution structure of a human histocompatibility molecule for her thesis and then stayed on to finish the work. She continued her postdoctoral training at Stanford University in the laboratory of Mark Davis before joining the staff at Caltech. Dr. Bjorkman received the William B. Coley Award for Distinguished Research in Fundamental Immunology from the Cancer Research Institute.

THE technique of x-ray crystallography allows visualization of the three-dimensional structures of proteins in atomic detail. In other words, we get a picture of the protein that shows the location of all the atoms and how they interact. The shape of a protein and the location of individual atoms with respect to one another are important for determining how the protein functions. With such knowledge it is often possible to design compounds that modify the protein for medical intervention.

The proteins that we are studying structurally are those that mediate the immune response against viruses and other pathogens. The immune system has evolved so that highly specific molecules on the surfaces of lymphocytes can recognize a virally infected cell. In the infected cell, pieces of viral proteins are fragmented and bound to a cellular protein called a histocompatibility molecule. If the complex formed between the histocompatibility molecule and the viral fragment is recognized by a protein on the lymphocyte, the infected cell is destroyed. The lymphocytes that bear the recognizing proteins are T cells, and the proteins on their surface are T cell receptors.

The three-dimensional structure of a histocompatibility protein gives us a picture of how and where viral molecules bind and how T cell receptors might bind to the complex formed between the viral and histocompatibility molecules. My laboratory now seeks to determine a three-dimensional structure for a T cell receptor, in order to understand the atomic details of its interaction with the surface of an infected cell. (This work is being done in collaboration with Mark Davis, HHMI, Stanford University.) We have crystallized a T cell receptor and now seek to make crystals of a complex between a T cell receptor and histocompatibility molecule. An understanding of how T cell receptors interact with histocompatibility molecules complexed to viral fragments should increase our understanding of how the immune system distinguishes normal, healthy cells from virally infected cells that need to be destroyed.

Our laboratory is also using protein expression systems in mammalian and bacterial cells to produce the large quantities of proteins necessary for crystallization and structural studies. Using molecular biological techniques, it is possible to transfect a protein-encoding gene into a cell in which it is not normally found, thus persuading that cell to manufacture that protein. Many proteins normally occur in small quantities; the use of such protein expression systems allows the isolation of much more of the protein than would be otherwise possible. We have expressed histocompatibility proteins in bacteria and formed complexes between these proteins and antigenic peptides.

We have also used a similar system to make an Fc receptor, a protein that binds to the Fc portion of antibody molecules. This particular Fc receptor is found in the intestine of newborn mammals and binds immunoglobulin found in mother's milk, thus transferring partial immunity from mother to infant. The amino acid sequence and structural organization of this molecule show similarities to histocompatibility molecules. We have made a soluble (non-membrane-bound) version of this molecule, which is still capable of binding to Fc molecules. We have recently purified and crystallized this Fc receptor and initiated a three-dimensional structure determination. We have also crystallized a complex between the receptor and the Fc portion of an antibody; this will ultimately allow us to obtain a picture of how the two proteins interact. A comparison of the Fc receptor structure and its interaction with antibodies should increase our understanding of the structurally related histocompatibility proteins and perhaps reveal reasons for the evolution of this type of structure in the immune system. We also have crystals of a structurally unrelated Fc receptor and are interested in comparing how the two types of receptors bind to their common ligand.

Mechanisms of Insulin Action

Perry J. Blackshear, M.D., D.Phil.—Investigator

Dr. Blackshear is also Professor of Medicine and Assistant Professor of Biochemistry at Duke University Medical Center. He received his D.Phil. degree in biochemistry from Trinity College, Oxford University, and his M.D. degree from Harvard Medical School. Before moving to Duke University, he was Assistant Professor of Medicine at Harvard Medical School. Dr. Blackshear has received the Young Investigator Award for Clinical Research from the American Federation for Clinical Research.

OUR laboratory is mainly interested in the molecular mechanisms of action of insulin and polypeptide growth factors. The studies involving insulin action are particularly relevant to common clinical disorders of insulin resistance, such as type II (adult-onset) diabetes and obesity, in which the locus of the resistance is thought to be at a "postreceptor" step within the cells of muscle, liver, and adipose tissue. Our work is aimed at understanding the biochemical steps involved in insulin action in these tissues, with the ultimate hope of identifying the abnormal steps in these insulin-resistant states and possibly using this knowledge to develop novel drugs aimed at correcting the abnormalities.

Another major area of interest is the molecular steps involved in mediating the effects of a wide variety of hormones, neurotransmitters, and drugs on their target cells. The common denominator of these agents is that the mechanism of action involves the stimulated breakdown of certain membrane lipid compounds, leading to the generation of intracellular lipid mediators known as diacylglycerols. These in turn can activate an important cellular enzyme, protein kinase C (PKC).

Over the past several years, a large amount of information has accumulated about the way in which hormones stimulate the breakdown of these membrane lipids and about the molecular biology and biochemistry of the PKC family of enzymes. However, almost nothing is yet known about the proteins that this kinase phosphorylates in cells and tissues, or about the involvement of these phosphorylated proteins in mediating a variety of cellular effects. To understand the role of these PKC substrates in the cell is a major goal of our laboratory. Within the past year we have made several advances in each of these areas.

With regard to signal transduction, we are investigating the molecular nature of a family of PKC substrates known by the acronym MARCKS (myristoylated alanine-rich C-kinase substrates). These are widely distributed cellular proteins that are phosphorylated within seconds of PKC activation in intact cells, and it is presumed that this phosphorylation serves some function in mediating the effects of the activated kinase. Our group has established that these proteins are excellent substrates for PKC, and that phosphorylation leads to changes in their properties that may influence their behavior in the cell. For example, phosphorylation of the protein disrupts its association with model cellular membranes, perhaps allowing for phosphorylation-dependent changes in the protein's intracellular location.

Similarly, phosphorylation of the protein disrupts its ability to form a tight complex with calmodulin, the ubiquitous calcium-dependent regulator of a wide variety of cellular enzymes. We postulate that such disruption leads to an increase in the cellular content of free calmodulin, which might then lead to concomitant activation of calmodulin-sensitive enzymes. In this way phosphorylation of the MARCKS protein and its relatives by PKC might lead to increases in the activity of calcium/calmodulin-dependent processes, a synergistic relationship that has long been postulated but not proved.

The overall goal of these studies is to try to determine the role that the MARCKS protein family plays, if any, in mediating the actions of PKC in various tissues. One important series of studies now under way involves attempts to render cells and intact animals deficient in these proteins, using the techniques of antibody microinjection, antisense RNA expression, and ultimately gene disruption. It is hoped that within the next year or two, such deficient cells and animals will be available to study the potential role of this family of proteins in development, particularly of the nervous system, in which the MARCKS proteins and their relatives are highly expressed.

Considerable progress has also been made in the studies of the molecular mechanisms of insulin action. In one group of studies involving two genes whose transcription is rapidly stimulated by insulin, we have made progress in identifying regions of the DNA that are implicated in the insulin reaction. We have also identified proteins that bind to these regions and have demonstrated that they are modified in some way by rapid insulin exposure of the cells. Current studies in the laboratory are attempting to clone and sequence

these binding proteins and to determine the nature of the insulin-stimulated modification.

Another group of studies involves the rapid insulin-stimulated translation of messenger RNA, using the enzyme ornithine decarboxylase in a model system. Within the past year, we have determined that insulin can rapidly stimulate the translation of certain types of mRNA that contain extremely extensive and convoluted secondary structures at their 5′-untranslated ends. We have proposed that insulin does this by means of a specific ability to activate translation initiation factors that can unwind or "melt" the RNA secondary structure, and have determined that insulin can rapidly stimulate the phosphorylation and presumably the activation of these initiation factors in intact cells.

Current studies are focusing on the protein kinases and phosphatases involved in this initiation factor phosphorylation, as well as on other rapidly stimulated, insulin-activated protein kinases that could play a role in the phosphorylation of the transcription factors alluded to above. In all of these cases, the ultimate goal is the elucidation of the biochemical steps between insulin binding to its receptor and its ultimate intracellular effect, such as gene transcription, with the hope that steps in this pathway that are abnormal in states of insulin resistance can be identified, leading to the development of specific therapies.

Intracellular Protein Traffic and Nuclear Organelles

Günter Blobel, M.D., Ph.D.—Investigator

Dr. Blobel is also Professor of Cell Biology at the Rockefeller University. He received his M.D. degree from the University of Tübingen and his Ph.D. degree with Van Potter in oncology from the McArdle Laboratory at the University of Wisconsin–Madison. Thereafter, he did postdoctoral work with George Palade at Rockefeller. Dr. Blobel is a member of the National Academy of Sciences and of several other distinguished societies. He has received many honors, including the Gairdner Foundation Award.

NUMEROUS structurally and functionally diverse proteins can be translocated across a few distinct cellular membranes. It is now established that targeting to these membranes and translocation across them is specified by a membrane-specific "signal" sequence that is a transient (or permanent) part of the protein to be translocated. The primary structure of numerous representatives for such a sequence has been established in our laboratory and by others. Present work focuses on the identification and characterization of machinery involved in the recognition of a signal sequence, in its targeting to the proper membrane, and in protein translocation. Collectively, these components comprise what has been termed a protein translocon.

Our laboratory is working on translocons for four distinct cellular entities: 1) the endoplasmic reticulum (ER) of animal (and yeast) cells, which is able to translocate proteins from the cytosol to the ER lumen; 2) bacterial plasma membranes (gram-negative bacteria), able to translocate proteins from the cytoplasm to the periplasmic space; 3) yeast mitochondria, able to translocate protein from the cytoplasm to the mitochondrial interior ("matrix") across outer and inner mitochondrial membranes; and 4) plant cell (pea and spinach) chloroplasts, able to translocate proteins from the cytoplasm to the chloroplast interior ("stroma") across outer and inner chloroplast membranes.

From studies on these four translocons so far, but especially from studies on the ER translocon, it is likely that a translocon is composed of at least four entities: 1) a soluble signal-recognition factor (SRF), 2) a homing receptor, 3) a signal sequence–gated protein-conducting channel, and 4) a signal-removing peptidase. The SRF has two functions: recognition of the signal sequence and targeting to the homing receptor, which is restricted in its localization to a translocon-specific membrane. Interaction of the signal sequence–SRF complex with the homing receptor leads to dissociation of the signal sequence and its presentation to a signal sequence receptor, which might be a subunit of the protein-conducting channel. This channel would close immediately following completion of translocation, only to open again after presentation of another signal sequence. The signal peptidase would, in most cases, remove the signal sequence either during or shortly after translocation.

In the case of the ER, the SRF was isolated and shown to be a ribonucleoprotein particle. Referred to as a signal-recognition particle (SRP), this consists of one 7S RNA molecule and six different proteins. Likewise, a homing receptor (referred to as the SRP receptor) and a signal-removing peptidase, a complex of five proteins, were isolated. More recently we were able to demonstrate the existence of a protein-conducting channel, using electrophysiological methods. The protein constituents of this channel remain to be identified. We have been able to solubilize the ER membranes by detergent and to reconstitute translocation-competent vesicles. Using this method, it should be possible to identify the channel proteins.

We recently identified components of other translocons. An SRF was isolated for signal sequence targeted to the bacterial plasma membrane. Moreover, we identified signal sequence–binding subunits for the mitochondrial and chloroplast translocons. These integral membrane proteins are located in contact sites between outer and inner organelle membranes and are candidates for subunits of a protein-conducting channel in the outer organelle membrane linked to a protein-conducting channel in the inner membrane.

Our other major research effort focuses on the organelles associated with the cell's nuclear envelope membranes. These organelles are thought to organize the large amount of information in the linear structure of the DNA into numerous structurally and functionally distinct three-dimensional superstructures, allowing only a limited amount of that information to be expressed. Characterization of these organelles should advance understanding of such fundamental processes as differential gene expression, cell differentiation, and development.

Our efforts focus on the structural and functional characterization of two morphologically

distinct structures. One, the nuclear pore complex (NPC), is located in the nuclear envelope. We speculate that this organelle is involved in "gene gating"; i.e., each pore complex is attached to a set of actively transcribed genes. We have identified and isolated several proteins of the NPC of mammalian cells. So far we have established the primary structure of 3 of the more than 100 proteins of this large organelle. Recently we succeeded in isolating NPCs from yeast. These genetically amenable cells should facilitate the structural and functional analysis of the many NPC proteins. We have also established the primary structure of two membrane proteins that are specifically located in the pore membrane domain of the nuclear envelope membrane. We are using cell-free systems for protein uptake into the nucleus to isolate the hypothetical SRF, the homing receptor, and the so-called transporter of the pore complex.

The other morphologically distinct structure associated with the nuclear envelope that we are studying is the nuclear lamina, a fibrous meshwork associated with the nuclear side of the inner nuclear envelope membrane. The lamina consists of three proteins, which we have termed lamins A, B, and C. We have speculated that the nuclear lamina is involved in the three-dimensional organization of nontranscribed chromatin.

By molecular cloning and cDNA sequencing of the three lamins, we and others showed recently that these proteins are members of the intermediate filament protein family. We demonstrated that lamin B binds to the carboxyl-terminal portion of cytoplasmic intermediate filament proteins and that ankyrin, a protein associated with the plasma membrane, binds to the amino-terminal portion of cytoplasmic intermediate filament proteins. Recently we have investigated the interaction of such a protein with nuclear lamin B in more detail and have localized the interacting regions to the near carboxyl-terminal segment of these proteins. Interestingly, a synthetic peptide, representing this lamin B–binding site of the cytoplasmic intermediate filament protein, when microinjected into cells, led to the detachment of intermediate filaments from the nucleus.

These data indicate a direct connection between the plasma membrane skeleton (ankyrin), the cytoskeleton (intermediate filaments), and the peripheral nucleoskeleton (lamina) through the NPC. Moreover, we recently identified a lamin B receptor in the inner nuclear membrane. Its primary structure has been determined by molecular cloning and cDNA sequencing. This receptor has a number of interesting sites, suggesting that it may not only interact with lamin B but also with DNA.

Immunity and Pathogenesis of Third World Diseases: Leprosy and Tuberculosis

Barry R. Bloom, Ph.D.—Investigator

Dr. Bloom is also Weinstock Professor of Microbiology and Immunology at Albert Einstein College of Medicine. He received his B.A. degree and an honorary Sc.D. degree from Amherst College and his Ph.D. degree from the Rockefeller University. He is active as an advisor to the World Health Organization in the areas of tropical diseases and vaccine development. Dr. Bloom also serves on the Board of Science and Technology in Development of the U.S. National Research Council and the National Vaccine Advisory Committee. He is a member of the National Academy of Sciences, the Institute of Medicine, and the American Academy of Arts and Sciences.

THE commitment of our laboratory is to investigate basic scientific problems that have particular relevance for health in the Third World. Three-quarters of the world's population lives in the Third World, and one-fourth of that population suffers from malnutrition and disease. The premises of our research are that the advances in molecular biology and immunology have a great deal to offer for understanding infectious diseases afflicting people in developing countries and, reciprocally, that the study of some of those diseases can provide insights into fundamental immunological and pathogenetic mechanisms of relevance to people in the industrialized world.

The Importance of Mycobacterial Diseases

Tuberculosis and leprosy are both caused by mycobacteria. Leprosy afflicts 4–6 million people in the world and produces deformity in 30 percent if untreated. Throughout time and in all cultures, leprosy has engendered a unique fear and stigma. Although *Mycobacterium leprae* was the first major human bacterial pathogen described, it remains one of the few that has never been cultivated in the test tube. However, its antigens can be produced and studied vicariously in *Escherichia coli* by means of recombinant DNA technology.

Tuberculosis is the major cause of death from a single infectious disease in the world today. Each year there are 8 million new cases of tuberculosis and 3 million deaths, afflicting primarily the most productive element of society—young adults. Infection with HIV (human immunodeficiency virus) causes a breakdown of resistance to tuberculosis; this has produced a grave increase in the disease, in both the developing and the industrialized countries. In 1985, following a 32-year decline in the number of cases, the incidence of tuberculosis in the Unites States began to increase, reaching 25,701 cases reported in 1990.

Resistance to *M. tuberculosis*

The devastatingly heightened susceptibility of HIV-infected individuals to a galloping course of tuberculosis is compelling evidence that there are powerful mechanisms of immunity in immunocompetent hosts. The cellular and molecular mechanisms mediating that immunity remain enigmatic. We have, however, established that the most common microbicidal product of activated macrophages, oxygen radicals, is ineffective at killing virulent human tubercle bacilli, but reactive nitrogen intermediates (particularly nitric oxide) can inhibit growth and kill *M. tuberculosis in vitro*.

Immunologic Unresponsiveness and Leprosy

One fundamental issue in immunology is the nature of immunological tolerance, i.e., the mechanisms by which cells in the immune system discriminate between foreign antigens and self-antigens and prevent responses to self. A breakdown of tolerance to self-antigens leads to autoimmune diseases, such as rheumatoid arthritis, juvenile diabetes, and perhaps multiple sclerosis. The principal mechanism for developing tolerance is thought to be deletion of clones of potentially autoreactive T cells in the thymus during neonatal life. Clearly, however, not all such clones can be deleted in the thymus; there must be additional mechanisms by which self-reactive cells can be rendered unresponsive after birth.

Leprosy provides a unique model with which to study immunoregulation and unresponsiveness. The disease comprises a spectrum of clinical entities. In the tuberculoid form, strong cell-mediated immunity kills the organism but damages nerves in the process. In the lepromatous form, at the other end of the spectrum, patients are unable to respond immunologically to *M. leprae* antigens. Because infection occurs after birth, there is little evidence of clonal deletion of T cells capable of reacting to this organism. Therefore understanding the mechanisms of that unresponsiveness is relevant to preventing

autoimmune disease and increasing transplant survival.

We have learned that the unresponsiveness in leprosy is antigen-specific: lepromatous leprosy patients unable to respond to antigens of *M. leprae* usually respond to those of *M. tuberculosis*, which is a closely related mycobacterium. How is it possible for T cells to recognize antigens in the tubercle bacillus and yet be unable to recognize the same or closely related antigens associated with the leprosy bacillus? We proposed that there might be one or a few unique antigens associated with *M. leprae* that induce active T cell suppression of potentially reactive T cell clones. Suppressor cells in immunology have been a controversial subject, but the idea that some T cells can down-regulate immune responses, particularly self-destructive ones, is compelling. About 85 percent of patients with lepromatous leprosy have a subset of T cells capable of being triggered specifically by leprosy antigens to suppress responses of immune T cells. Although they represent a minor subset of T cells in the blood, they are the major lymphocyte subset in lepromatous lesions. By establishing long-term T cell clones directly from the lesions and blood, we found that the suppressor cells have a pattern of antigen recognition different from other cytotoxic or lymphokine-producing T cells. They carry the surface marker CD8 and recognize foreign antigens in association with the HLA-DQ region of the human major histocompatibility complex (MHC) class II. We speculate that presentation of antigens by this MHC subset predisposes the immune responses toward negative rather than positive responses. Our studies suggest that functionally distinct subsets of T cells are characterized by distinct patterns of lymphokines produced upon antigenic stimulation.

New Vaccines from Old—Recombinant BCG as a Multivaccine Vehicle

Vaccines represent the most cost-effective medical intervention. Yet three general problems limit the use of many current vaccines: 1) they require multiple booster shots to be effective; 2) they cannot be given for 6–12 months after birth, because of transferred maternal antibodies that inactivate them; and 3) the cost. BCG (bacille Calmette-Guérin), the most widely used vaccine in the world, is a live, attenuated bovine tubercle bacillus given to protect children against tuberculosis.

BCG has been given to more than 2.5 billion people and has a very low incidence of serious side effects. It is one of only two childhood vaccines that can be given at birth or any time thereafter. It is a single-shot vaccine that engenders long-lasting cellular immunity and costs only $0.10 a dose.

The unique attributes of BCG suggested to us that, if it could be genetically engineered to express a variety of foreign antigens protective for different pathogens, a single immunization might be capable of engendering protective responses to multiple pathogens simultaneously. One problem, however, was the paucity of molecular genetic information about the *Mycobacteria*. In collaboration with William Jacobs (HHMI, Albert Einstein College of Medicine), we developed genetic systems for introducing and expressing foreign genes in mycobacteria, particularly BCG strains. We developed a shuttle strategy in which mycobacterial DNA could be genetically cloned and manipulated in *E. coli* and then transferred into mycobacteria. Our first approach was to use mycobacteriophages (viruses that infect bacteria) as vectors to target foreign genes to a specific site in the bacterial chromosome. This enabled us to introduce foreign DNA into BCG for the first time. Recently we have developed shuttle plasmid vectors that are able to produce many copies of foreign genes in BCG.

With several collaborators at MedImmune, Inc., and the University of Pittsburgh, we have developed the first experimental recombinant BCG vaccines. These express protective antigens from *M. leprae*, schistosomes, malaria, measles virus, leishmania, and HIV. Initial experiments in mice indicate that three major types of protective immune responses can be generated *in vivo*—namely immunoglobulin antibodies, T cell lymphokines, and cytotoxic T lymphocytes. Continuing efforts will be made to define antigens that will engender, through recombinant BCG, protective immunity against a variety of viral, bacterial, and parasitic pathogens.

Molecular Biology of the Extracellular Matrix

Jeffrey F. Bonadio, M.D.—Assistant Investigator

Dr. Bonadio is also Assistant Professor of Pathology and Member of the Program in Bioengineering at the University of Michigan Medical School. He received his bachelor's degree in biology from Marquette University and his M.D. degree from the Medical College of Wisconsin, Milwaukee. He studied anatomical pathology with Bruce Beckwith and medical molecular genetics with Peter Byers at the University of Washington, Seattle.

THE long-term goal of our research is to understand how extracellular matrix proteins contribute to skeletal structure and function. Quantitative and qualitative changes in these proteins occur during morphogenesis and as part of the wound healing process. These observations suggest that both the organization and protein composition of the matrix are precisely regulated. It is clear that this regulation occurs in part at the level of gene expression and in part at the level of the assembly of proteins into a matrix-like configuration.

I have chosen to focus for the most part on the matrix molecule type I collagen. This collagen is a polymer of two related proteins whose sequence has been determined. Moreover, the multidomain structure of the molecule and a general outline of collagen biosynthesis are known, and the molecule is recognized to be distributed widely within tissues such as bone, tendon, ligament, tooth, dermis, and sclera. Previous studies have implied that type I collagen makes an important contribution to the structure, integrity, and normal homeostasis of these tissues. Over the past year we have continued our work to establish model systems that would allow us to study this contribution at the molecular level.

One system is designed to investigate the intracellular assembly of the collagen molecule. In general, this work involves site-specific mutagenesis and assays that quantify the effects of mutation on the assembly process. These effects are studied at two levels. First, we have established conditions that allow synthetic peptides to fold into a collagen-like triple helix. Peptide folding is slow enough that the process can be characterized by methods such as circular dichroism. In addition, the triple helix formed *in vitro* is stable enough that its structure can be characterized by nuclear magnetic resonance (NMR) techniques. Therefore the effect of a given mutation can be quantified by directly comparing the behavior of a normal peptide with that of mutant peptide. Second, cellular transfection methods have been developed to express and assemble collagen molecules *in vitro*. Again, the effect of mutation on the assembly process can be quantified by directly comparing the behavior of normal and mutant molecules.

In our initial mutagenesis experiments, we characterized a highly conserved region of the triple-helical domain and demonstrated that it made an important contribution to the assembly of collagen molecules into a thermodynamically stable conformation. We speculate that this region was conserved during evolution because it plays an important role in collagen biosynthesis, i.e., in folding the collagen molecule into its correct conformation. In the future, we hope to use this model system to define further the normal contribution made by other collagen domains to the assembly process. In addition, we are interested in characterizing those regions of the molecule that mediate interactions between collagen and other matrix molecules such as fibronectin, heparin sulfate proteoglycan, and integrins. These interactions are important because they represent a molecular basis for the assembly of collagen within the matrix.

A second system is designed to investigate the function of type I collagen at the level of connective tissue. Our initial set of experiments utilized a transgenic mouse strain that expressed only half the normal amount of type I collagen. We demonstrated that the mutation adversely affected the connective tissue of bone and skin dermis. In addition, the mutant mice were profoundly deaf. We utilized biomechanical tests to quantify the effect of the collagen deficiency at the tissue level, and these studies demonstrated that the major role of type I collagen is to provide connective tissue with a high degree of resiliency.

More recently, we also demonstrated that the skeleton of these transgenic mice is able to adapt to the inherited collagen deficiency. This adaptation involves a thickening of cortical bone and results from the synthesis of new bone matrix. Particularly intriguing was our observation that the adaptation was associated with a significant improvement in bone strength. This result is important because it suggests the basis for a strategy to strengthen the fragile skeleton.

Functional Heterogeneity in CD4-bearing T Lymphocytes

H. Kim Bottomly, Ph.D.—Associate Investigator

Dr. Bottomly is also Associate Professor in the Section of Immunobiology and the Department of Biology at Yale University School of Medicine. She received her Ph.D. degree from the University of Washington, Department of Biological Structure, where she studied with Roy Schwartz. Her postdoctoral training was received in the field of immunology with Don Mosier at NIH.

LYMPHOCYTE interactions during an immune response are necessary for the induction of antigen-specific lymphocytes. The resulting effector phase of an immune response is described as either humoral or cell mediated: each of these phases combats different types of microorganisms. Although these effector mechanisms are well characterized, the precise mechanism by which the response to a given antigen or infectious agent is directed into the humoral or cell-mediated mode is not known. What is clear, however, is that both types of immunity depend on the activation of CD4-bearing T lymphocytes, which in turn induce other cell types to respond to the foreign antigen. These responses include activation of B cells to proliferate and to secrete antibody, induction of delayed-type hypersensitivity reactions, activation of CD8 cytolytic T cells, and activation of macrophages.

Two questions were then asked. 1) Could the same CD4 T cell activate all these target cells, therefore mediating both humoral and cell-mediated responses? 2) Is the CD4 T cell population functionally heterogeneous; i.e., do some CD4-bearing T cells activate B cells and play a primary role in the induction of humoral immunity, and do some activate macrophages and CD8-bearing T cells and play a primary role in cell-mediated immunity?

For several years we have focused on the heterogeneity in CD4 T cell function and the activation conditions that lead to it. These studies have successfully shown that monoclonal CD4 T lymphocytes, obtained by T cell cloning and expansion in tissue culture, belong to two distinct subsets of T cell. One subset can help B cell antibody secretion but cannot activate macrophages; the other set activates macrophages but is a poor B cell activator. To reflect their main functions, these two subsets are called helper T cells (Th2) and inflammatory T cells (Th1). The distinct functional abilities of Th2 and Th1 subsets are reflected in their release, upon activation, of distinct panels of cytokines. Several cytokines are produced selectively by one or the other subset. In particular, interleukin-2, interferon-γ, and lymphotoxin are produced by the Th1 but not Th2 cell subset. Interleukin-4 is produced by the Th2 but not Th1 cell subset.

Thus there is a correlation between cytokine production and function. Interleukin-4 is a potent B cell activator involved in B cell proliferation and secretion of immunoglobulins IgG1 and IgE. By contrast, interferon-γ, lymphotoxin, and interleukin-2 are associated with responses involving the activation of macrophages, lysis of target cells, and induction of cytolytic T cells, which is consistent with the known function of these cytokines. One can conclude from these studies that T cells are committed to the release of a distinct panel of lymphokines when activated, with the released effector molecules determining their effector function.

Recent studies in this laboratory have focused on determining whether functionally distinct subsets of CD4 T cell exist *in vivo* and whether the selective activation of one subset or the other has the expected functional consequences. Numerous clinical studies have suggested that various immunization schemes induce primarily a humoral or cell-mediated immune response. It is critically important to determine if such major differences in protective immunity reflect differences in the proportion of CD4 T cell subsets activated. If this is true, one might propose that the form of the antigen or the antigen presentation must somehow direct which CD4 T cell will be preferentially activated. To test this possibility, several questions about normal CD4 T cells have been asked.

First, is there a separation of resting CD4 T cells into subsets? To answer this, analysis of the responses to antigens that give rise to primarily humoral or cell-mediated immunity has been performed. When humoral immunity dominates, Th2-like CD4 T cells are activated. By contrast, when cell-mediated immunity dominates, Th1-like CD4 T cells are activated. Thus during the course of an immune response, CD4 T cells may become specialized in their functional capabilities, and these T cells *in vivo* resemble, in their activities, Th1 and Th2 cloned lines.

Second, when during development or activation do subsets of CD4-bearing T cells arise? A particularly interesting question is whether the commitment of CD4-bearing T cells to the Th1 and Th2 subsets is a consequence of antigen priming. Analysis of memory and virgin CD4-bearing T cells indicates that the functional specialization characteristic of Th1 and Th2 subsets resides mainly in the memory population. This suggests that contact with foreign antigen initiates a commitment of CD4-bearing T cells to a particular effector response.

The form an immune response takes must be appropriate to the microorganism that is causing the disease. This is seen clearly in leprosy, where one response produces abundant antibody and, at the same time, overwhelming growth of the leprosy bacterium in macrophages. This is probably an example of selective activation of the helper (Th2), rather than the inflammatory (Th1), subset of CD4 T cells. What controls the form such a response will take? Model systems to explore these questions are needed, so that an inappropriate immune response can be redirected to become curative. Future research in this laboratory will focus on this question of the control of CD4 T cell subset activation and regulation.

Studies analyzing those factors important in the generation of memory and effector $CD4^+$ T lymphocytes and those analyzing the selective activation of CD4 T cell subsets are supported by grants from the National Institutes of Health.

Retroviral Replication

Patrick O. Brown, M.D., Ph.D.—Assistant Investigator

Dr. Brown is also Assistant Professor of Pediatrics and of Biochemistry at Stanford University School of Medicine. He received his B.A. degree in chemistry from the University of Chicago. His graduate work with Nicholas Cozzarelli at the University of Chicago was focused on the mechanisms of DNA topoisomerases. He received his Ph.D. and M.D. degrees from the University of Chicago, completed a pediatrics residency at Children's Memorial Hospital in Chicago, and then joined Michael Bishop's laboratory at the University of California, San Francisco. There he began to investigate the mechanism of retroviral integration, which has continued to be the major focus of his research.

RETROVIRUSES are an important cause of disease in most vertebrate species. In humans, retroviral infections can lead to AIDS (acquired immune deficiency syndrome), leukemia, lymphoma, and degenerative diseases of the central nervous system. Millions of people are infected with the human immunodeficiency virus, HIV, and will likely succumb to AIDS unless an effective treatment is developed.

The retroviral genes are carried in the virus particle as RNA molecules. When the virus infects a cell, it transcribes these molecules, its RNA genome, into a double-stranded DNA molecule and inserts this into a chromosome in the nucleus of the host cell. Thus the viral genome, then called a provirus, becomes an integral part of the cell's genome. Integration of a provirus into its host cell's DNA is essential for retroviral reproduction. This distinctive feature of the retroviral life cycle accounts for many of the characteristics associated with retroviral infection, including insertional mutagenesis, induction of tumors, and the latent and persistent nature of many retroviral infections. Moreover, the fact that retroviruses are designed to introduce foreign genes into cellular DNA makes them exceptionally useful as tools for genetic engineering.

How does a retrovirus get its DNA into a cell's nucleus and integrate it into the cell's DNA, and how does the cell regulate these processes?

To investigate the molecular mechanism by which a retrovirus inserts its DNA into that of the infected cell, we have developed a variety of methods for studying the retroviral integration reaction in a test tube. We have used this approach to define several discrete steps in the joining of viral to cellular DNA and to determine the requirements for the reaction. The enzymatic machinery that carries out integration can be isolated from infected cells in a stable complex with the unintegrated viral DNA molecule. We are currently using electron microscopy as well as biochemical methods to define the structure of this viral replication intermediate.

To study the enzymology of integrase, the viral protein that actually catalyzes integration, we have constructed genetically engineered bacterial strains that produce abundant quantities of the integrase proteins from HIV and murine leukemia virus and have developed simple purifications of these proteins. Using small synthetic DNA molecules as model substrates, we can now readily study their catalytic activities, which include the sequence-specific processing of the ends of the viral DNA and the joining of these ends to a target DNA molecule. We have made progress in the past year toward understanding the organization of integrase and how it binds the viral and target DNA substrates. For example, we have identified DNA substrates that function preferentially either as donors or as targets in a DNA joining assay. By studying the competition of these DNAs for binding to the enzyme, we have identified two distinct sites that bind viral and target DNA, respectively. By analyzing the reversal of the usual DNA joining reaction, we have discovered that integrase has a previously unrecognized DNA splicing activity. We are investigating the possibility that this new activity may play a role in integration and in the high-frequency recombination that occurs between viral genomes during replication.

The structure of integrase is clearly of central importance to understanding integration. Because efforts to obtain conventional crystals of integrase have been unrewarding, we are preparing to use electron diffraction methods to determine a structure from two-dimensional crystals. By constructing specific mutants and defining their biochemical defects, we have begun to identify the functions of specific protein regions. To extend this approach, we are developing a new genetic system that we hope will enable us to screen for functional defects among millions of mutant integrases, based on their ability to carry out recombination in bacteria. This system should also facilitate our efforts to develop genetically altered integrases with properties more favorable for therapeutic applications. For example, we would like to develop an integrase that

can selectively integrate the viral DNA into predetermined sites in the target DNA.

The ultimate goal of our work on integration is to understand in molecular detail how the proteins of the integration machinery recognize the viral DNA, assemble into an active complex, recognize the target DNA, and finally catalyze the DNA breakage and joining reactions that lead to integration of the provirus. It is hoped that this understanding will lead to the development of new agents for inhibiting the replication of pathogenic retroviruses and to improved systems for the therapeutic introduction of genes into mammalian cells.

It has been recognized for many years that establishment of a retroviral provirus proceeds much more readily in actively dividing cells than in their resting counterparts. However, the basis for this phenomenon remains obscure. To bring this observation into clearer focus, we have investigated the dependence of specific steps in the life cycle of the murine leukemia virus on the host cell's stage in its own division cycle. Using drugs or a mutation in the cell-cycle control gene *cdc2* to regulate the cell cycle of the host cells, we have found that integration *in vivo* depends on mitosis. Yet synthesis of viral DNA and accumulation of integration-competent intermediates occur normally in cells blocked from entering mitosis. The intermediates accumulate in the cytoplasm and remain stable for hours, awaiting mitosis. Current investigations are aimed at understanding why mitosis is required for integration. Preliminary evidence suggests that disassembly of the nuclear envelope at mitosis may provide a route of entry for viral replication intermediates. Understanding how cellular functions can determine the fate of an infecting retrovirus may lead to new approaches to antiviral therapy and to improvements in the use of retroviruses as vectors for gene therapy.

Our work on retroviral integration is supported by a grant from the National Institutes of Health.

New Methods for Linkage Mapping in Complex Genomes

A major impediment to defining and characterizing the genes that influence complex human traits has been the difficulty of collecting suitable large families in which the trait segregates. Such families are generally needed to find genes by conventional linkage mapping. The same genes could in principle be mapped more easily by an alternative strategy that involves collecting and analyzing pairs of relatives that share a trait of interest. However, linkage mapping with small sets of relatives generally requires analysis of a large number of closely spaced and highly polymorphic genetic markers, which makes this strategy impractical with current technology.

We are developing a new set of genetic tools that will allow widespread application of these highly efficient linkage mapping methods. Experiments are in progress to test these new methods in model systems. In parallel with our efforts to develop new biochemical methods, we are working to find optimal statistical methods, using high-resolution maps of genetic identity between pairs of relatives. Our aim is to apply this technology to map genes for complex human traits.

This work is supported by a grant from the National Institutes of Health.

Regulation of Cellular Processes by Protein-Tyrosine Phosphorylation

Joan S. Brugge, Ph.D.—Investigator

Dr. Brugge is also Professor of Microbiology at the University of Pennsylvania School of Medicine. She received her B.A. degree in biology from Northwestern University and her Ph.D. degree in virology from Baylor College of Medicine. Her postdoctoral research was done with Raymond Erikson at the University of Colorado School of Medicine, Denver. Before moving to the University of Pennsylvania, Dr. Brugge was a member of the Department of Microbiology at the State University of New York at Stony Brook.

THE modification of intracellular proteins by the reversible addition of phosphate groups (phosphorylation/dephosphorylation) is the most common mechanism for regulating their activity. Enzymes that specifically transfer phosphate to tyrosine residues of proteins (protein-tyrosine kinases) play critical roles in intracellular signal transduction events—biochemical processes that transform interactions at the surface of the cell into an intracellular response. Our laboratory is interested in investigating the role of tyrosine phosphorylation in these processes. We have chosen for these studies two model systems that offer unique advantages: platelets, which are ideally suited for studies of cell adhesion, secretion, and cytoskeletal rearrangements, and PC12 cells, which allow dissection of events that are involved in the differentiation of neuronal cells after treatment with nerve growth factors.

Tyrosine Phosphorylation in Platelets

Platelets are small, anucleate peripheral blood cells that contain many intracellular vesicles whose components are released upon activation by cellular hormones. Platelet aggregation and the released products from platelets are responsible for the formation of blood clots and wound healing. We have found that hormones that activate platelet aggregation and secretion cause rapid changes in the phosphorylation of multiple proteins on tyrosine. In collaboration with Sanford Shattil (University of Pennsylvania School of Medicine), we have shown that the phosphorylation of several of these proteins requires platelet aggregation.

This process is mediated by interactions between the serum adhesion protein fibrinogen and its receptor, GP IIb-IIIa, on the platelet surface. GP IIb-IIIa is a member of the integrin family of receptors that bind to extracellular matrix and adhesion proteins. These receptors are believed to be important in causing changes in cell behavior that are mediated by cell adhesion. Our studies demonstrating that inhibition of GP IIb-IIIa–induced cell aggregation prevents tyrosine phosphorylation of cellular proteins suggest that this receptor may regulate the activation of tyrosine kinases, which play a role in aggregation-dependent events.

We have recently found that a cytoplasmic tyrosine protein kinase, p125-FAK, is phosphorylated on tyrosine and activated following platelet aggregation. This kinase was first identified by Tom Parsons (University of Virginia School of Medicine) as a substrate phosphorylated in Rous sarcoma virus–transformed cells. The enzyme is localized in focal adhesion plaques—sites where integrin receptors couple with extracellular matrix proteins and intracellular microfilaments. It is not phosphorylated on tyrosine or activated in platelets from patients with Glanzmann's thrombocytopenia, a disorder caused by the absence of a functional GP IIb-IIIa. These results suggest that p125-FAK may be activated by events following GP IIb-IIIa–dependent platelet aggregation. It is also phosphorylated on tyrosine in fibroblasts following activation of other integrin receptors and could play a critical role in integrin-mediated activation of intracellular signal transduction events mediated by cell-cell and cell-matrix interactions.

GP IIb-IIIa may also be important for the assembly of intracellular cytoskeletal complexes that follow thrombin-induced platelet aggregation. Together with Joan Fox (Gladstone Foundation), we have found that the assembly of these cytoskeletal proteins with enzymes involved in signal transduction is dependent on GP IIb-IIIa and platelet aggregation. We have found that four members of the Src family of protein-tyrosine kinases—Src, Fyn, Lyn, and Yes—are part of these assemblies, as are the cytoskeletal proteins actin, vinculin, talin, and GPIV. In platelets from Glanzmann's thrombocytopenia patients, these proteins do not associate with the detergent-insoluble cell fraction, indicating that GP IIb-IIIa is critical for these cytoskeletal rearrangements. Thus platelet aggregation mediated by fibrinogen binding to GP IIb-IIIa appears to play an important role in nucleating the assembly of cytoskeletal signaling complexes that may be important for activation of intracellular processes, and these

studies raise the possibility that integrin receptors in other cell types play similar roles in coupling with cytoplasmic tyrosine kinases.

We have also found that another platelet membrane receptor, GPIV, or CD36, is tightly coupled with three other protein-tyrosine kinases. These kinases, Fyn, Lyn, and Yes, are members of the Src family of protein-tyrosine kinases, which have been shown to be linked with transmembrane receptors in other cell types. The ligand that binds to GPIV/CD36 has not been identified; however, the evidence that this receptor is associated with protein-tyrosine kinases strongly implicates it is an intracellular signal transducer.

The Fc receptors for the complement-binding domain of antibody molecules are also found on platelets. Activation of these receptors causes platelet aggregation and secretion, and the induction of tyrosine phosphorylation of the same proteins that are phosphorylated in thrombin-treated platelets. However, unlike other platelet receptors, the Fc-γRII receptor is itself phosphorylated on tyrosine. The site(s) of tyrosine phosphorylation of this receptor lies within a peptide motif that is shared with several other lymphocyte receptors in T cells, B cells, natural killer cells, mast cells, and basophils. This motif appears to be involved in coupling these receptors with protein-tyrosine kinases. Tyrosine phosphorylation of the conserved tyrosine residues in this motif may be important for these coupling interactions.

Tyrosine Phosphorylation in Neuronal PC12 Cells

We are interested in defining the role of tyrosine phosphorylation in mediating neuronal differentiation. PC12 cells, derived from a rat pheochromocytoma, provide a useful model system, since these cells differentiate into cells resembling sympathetic neurons after treatment with nerve growth factor (NGF) or fibroblast growth factor (FGF). Both of these receptors are transmembrane proteins whose cytoplasmic domains contain protein-tyrosine kinase activity. Growth factor binding to these cells causes a burst of tyrosine phosphorylation of multiple cellular proteins.

Recently, in collaboration with Simon Halegoua, we have found that tyrosine phosphorylation of two serine/threonine protein kinases, $p42^{MAPK}$ and $p44^{MAPK}$, is dependent on the small GTP-binding protein Ras. Expression of a mutant of Ras that interferes with the activity of the endogenous Ras protein blocks tyrosine phosphorylation of these proteins, and expression of a constitutively activated form of Ras leads to their tyrosine phosphorylation. These mutant forms of Ras have no effect on the activity of the NGF receptor itself or on the phosphorylation of proteins that are direct substrates of this kinase.

The p42 and p44 MAP kinases are referred to as "switch kinases." They are activated by phosphorylation on tyrosine but phosphorylate other proteins on serine and threonine. MAP kinases are activated by a variety of mitogens and growth factors, and are believed to serve as critical components of intracellular signaling pathways by integrating signals from a diverse array of receptors. Our studies suggest that tyrosine phosphorylation of MAP kinase is dependent on pathways regulated by the Ras GTP-binding protein, and that this event can thus be distinguished from phosphorylation events that are directly mediated by the NGF receptor itself. Since Ras activity is essential for NGF- and FGF-induced neuronal differentiation of PC12 cells, these results raise the possibility that MAP kinases are also essential components of this process.

Dr. Brugge is now Scientific Director at Ariad Pharmaceuticals, Cambridge, Massachusetts.

Computational Structural Biology

Axel T. Brünger, Ph.D.—Assistant Investigator

Dr. Brünger is also Associate Professor of Molecular Biophysics and Biochemistry at Yale University. He was born in Leipzig, Germany. He received his diploma in physics at the University of Hamburg and his Ph.D. degree from the Technical University of Munich. He held a NATO postdoctoral fellowship and subsequently became a research associate with Martin Karplus in the Department of Chemistry at Harvard University before joining the faculty at Yale. His research has focused on molecular dynamics studies of protein structure and function and on methods in protein crystallography and nuclear magnetic resonance spectroscopy.

OUR research lies at the interface between theory and experiment in the area of structural biophysics. The research tools are simulation methods of computational chemistry adapted to the requirements of macromolecular systems. Macromolecular simulations are an important addition to the arsenal of methods available to structural biologists working with x-ray crystallographic or nuclear magnetic resonance (NMR) spectroscopic data. In one class of projects, we are trying to understand the detailed microscopic interactions that govern stability and recognition in biological systems and to test the reliability of the theoretical methods as tools for this purpose. In another class of projects, we are directly combining macromolecular simulation with experimental data in order to make data analysis possible or more efficient.

Accuracy of Crystal and Solution NMR Structures

As methods for determining macromolecular three-dimensional structure continue to become more powerful and are being applied to many biologically interesting systems, concern has been raised about the verification of final atomic models. A common problem arises when models are fitted against preliminary experimental data of mediocre quality. The recent revision of a number of published structures, both x-ray and solution NMR, illustrates the need to develop improved tools for checking the accuracy of the final atomic models.

Structure determination of macromolecules by crystallography involves fitting atomic models to the observed diffraction data. The traditional measure of the quality of this fit, and presumably the accuracy of the model, is the R value. Despite stereochemical restraints, it is possible to overfit or "misfit" the diffraction data: an incorrect model can be refined to fairly good R values, as several recent examples have shown. We recently proposed a reliable and unbiased indicator of the accuracy of such models.

In analogy to testing statistical models by cross-validation, we defined a statistical quantity, R^T_{free}, that measures the agreement between observed and computed structure factor amplitudes for a "test" set of reflections that is omitted in the modeling and refinement process. As examples show, there is a high correlation between R^T_{free} and the accuracy of the atomic model phases. This is useful, since experimental phase information is usually inaccurate, incomplete, or unavailable.

The enhanced sensitivity of R^T_{free} with respect to model errors was illustrated for the crystal structure of the plant ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) (David Eisenberg, University of California, Los Angeles). A partially incorrect model for RuBisCO, which essentially had the small subunit traced backward, showed only a 4 percent conventional R value difference from the correct model. On the other hand, R^T_{free} showed a 13 percent difference, suggesting that the incorrect model had been overfit.

We concluded that R^T_{free} represents a reliable and unbiased parameter by which to evaluate the information content of a model produced by x-ray crystallography. It is not restricted to high-resolution diffraction data. The observation that R^T_{free} can distinguish between a random distribution of scatterers and a distribution close to the protein suggests applications to *ab initio* phasing. Presently we shall apply this method to assess models of thermal motion and disorder, time-averaging, and bulk solvent in protein crystals.

A similar approach might be useful for the three-dimensional structure determination by solution NMR. Using molecular dynamics refinement, we have just succeeded in implementing the free R approach. At present we are testing it for a number of model systems. The question of accuracy is an even more fundamental problem for solution NMR structures because of the adverse observable-to-parameter ratio.

This work is also supported by the National Science Foundation.

Predictions of Helix-Helix Association and Stability

Prediction of the three-dimensional structure

of proteins based on their sequence remains impossible. This fundamental problem of structural biology (the "folding problem") is still unsolved, despite improvements in computational techniques for macromolecular simulation and computer hardware. Nevertheless, macromolecular simulation has been successful in predicting localized conformations if sufficient experimental constraints or restraints (e.g., in the form of an x-ray structure) are available. It is therefore conceivable that other more global predictions are possible if appropriate experimental information is available. We have embarked on trying to predict the association and stability of helices that form coiled coils. Conformational search strategies are being employed with empirical energy functions, using molecular dynamics and energy minimization.

Presently we shall apply this approach to the family of leucine zipper proteins, which are sequence-specific DNA-binding proteins that regulate gene expression in certain mammalian cells. We have successfully predicted the structure of the dimerization domain of GCN4, for which a high-resolution x-ray has become available (Thomas Alber, University of Utah).

Macromolecular Simulation of Free-Energy Differences

We are involved in a number of projects aimed at simulating free-energy differences between two states of a biological system, using the so-called free-energy perturbation technique. The goal is to investigate microscopically the structure and stability of protein secondary structural elements and protein-peptide complexes. Furthermore, we would like to evaluate the reliability of free-energy calculations and molecular dynamics simulations as tools for this purpose.

One project concerns the complexes of bovine pancreatic ribonuclease S and a number of mutants of the S-peptide for which x-ray crystal structures, binding free energies, and enthalpies have been obtained by Frederic Richards and Julian Sturtevant (Yale University). Another project involves the study of a number of site-directed mutants of β-turns in staphylococcal nuclease in a joint project with Robert Fox (HHMI, Yale University) for which x-ray crystal structures and information about the *cis* to *trans* equilibria of a proline side chain in the turn have been obtained in Dr. Fox's laboratory.

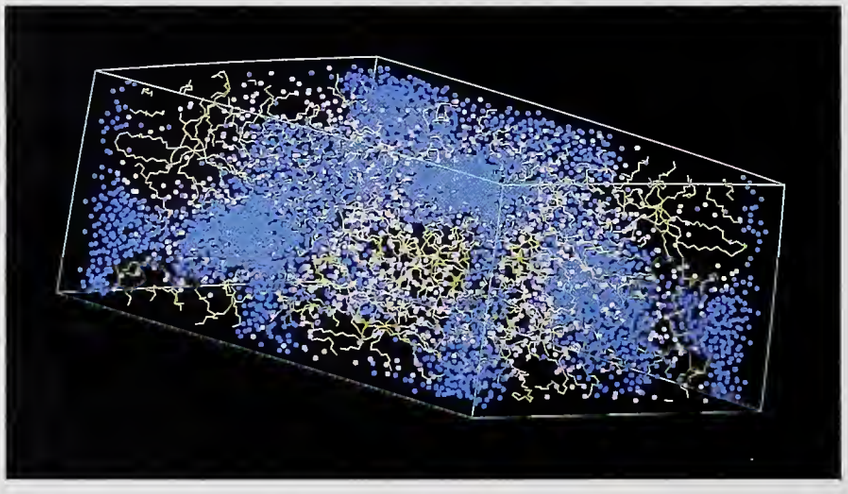


Illustration of bulk solvent regions in protein crystal structures. The regions are indicated by blue dots in this unit cell of a crystal of penicillopepsin from Penicillium janthinellum, *whose structure was solved in Michael James' laboratory (Alberta, Canada) at 1.8-Å resolution. The protein backbone atoms appear as yellow lines; ordered water molecules, as pink dots. Bulk solvent typically constitutes 40–60 percent of the unit cell contents, yet its physical characteristics and biological function are poorly understood.*

Research and photograph by Axel Brünger, using a graphics interface developed by Warren Delano.

Biophysical Studies of Eukaryotic Gene Regulation and Molecular Recognition

Stephen K. Burley, M.D., D.Phil.—Assistant Investigator

Dr. Burley is also Assistant Professor and Co-Head of the Laboratory of Molecular Biophysics at the Rockefeller University. He received a B.Sc. degree in physics from the University of Western Ontario, a D.Phil. degree in molecular biophysics from Oxford University, and an M.D. degree from Harvard Medical School in the Harvard-MIT Joint Program in Health Sciences and Technology. While a medical student, he carried out research in protein crystallography with Gregory Petsko. During his clinical training at Brigham and Women's Hospital, he also conducted postdoctoral research in protein crystallography with William Lipscomb at Harvard University, where he solved the three-dimensional structure of leucine aminopeptidase.

WE are interested in developing a detailed understanding of the physical principles that govern the general problem of molecular recognition in biological systems. Our approach is to use x-ray crystallography and complementary biophysical methods to determine and characterize the three-dimensional structure and function of biological macromolecules and their complexes with DNA, proteins, or smaller ligands. These structures contain a wealth of atomic detail that we can analyze with biochemical, molecular genetic, and theoretical methods to provide a functional description of the intra- and intermolecular interactions responsible for stabilizing macromolecular complexes.

In the long term, we hope that our biophysical studies and analyses will allow us to exploit the powerful formalism of physics to classify systematically the interactions between individual atoms that are responsible for molecular recognition in biological systems. We believe that such a quantitative understanding will ultimately permit us to harness the machinery of molecular recognition and, thereby, make defined interventions into important biochemical processes such as disease states.

Eukaryotic Gene Regulation

We are examining the problem of eukaryotic gene regulation, with the goal of improving our understanding of the structural and physical bases of transcriptional control of genes. Three distinct classes of proteins are active in transcription, and we are studying representative members of each class. First, we are collaborating with Robert Roeder (Rockefeller University) on x-ray crystallographic and complementary biophysical studies of transcription factor IID (TFIID) and other components of the basic transcription machinery. These proteins form a stable multiprotein, or preinitiation, complex with DNA sequences found immediately upstream of the transcription start site, where they mediate transcription by RNA polymerase II. We have purified, characterized, and crystallized TFIID, which begins preinitiation complex formation by binding to the TATA consensus sequence. In addition, we have started work on TFIIB, the second protein to be recruited to the preinitiation complex.

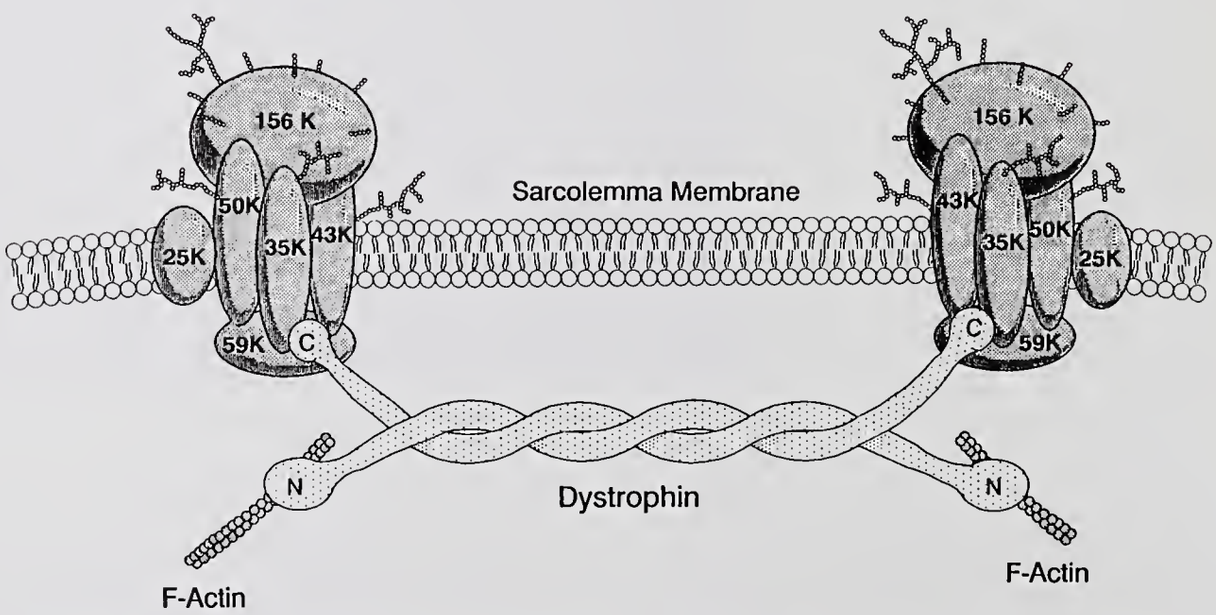
Second, we are also collaborating with Dr. Roeder on studies of upstream stimulatory factor (USF), a member of the c-*myc*–related family of DNA-binding proteins that contains both a helix-loop-helix motif and a leucine repeat. We have used various biophysical methods to purify and extensively characterize USF and its mechanisms of action. Moreover, we were recently able to grow small cocrystals of USF and DNA. During transcription, TFIID, the other basic factors, and USF bind to DNA in close proximity and interact with one another to enhance both DNA binding and transcription. After determining the three-dimensional structures of each of these proteins and their complexes with their respective promoter DNA elements, we hope to determine the structures of some biologically relevant multiprotein-DNA complexes.

Third, we are collaborating with Eseng Lai (Memorial Sloan-Kettering Cancer Center) on structural studies of human hepatocyte nuclear factor 3. This transcriptionally active protein belongs to a gene family in mammals that is homologous to the *Drosophila* homeotic gene *fork head*. These diverse proteins share a highly conserved DNA-binding region and influence transcription by binding to DNA elements, known as enhancer sequences, that are located far from the transcription start site. During the past year, we have cocrystallized a member of this family with DNA and are proceeding with a three-dimensional structure determination. Detailed structural and biophysical studies of these three distinct classes of participants in eukaryotic transcription should provide insights into the precise role molecular recognition plays in gene regulation.

Molecular Recognition

During the past year, we have also begun to study other biological systems that function

via molecular recognition, including steroid-binding proteins, polypeptide-binding proteins, and enzymes. Three-dimensional structure determinations of these proteins will be pursued once they are fully characterized and suitable crystals and cocrystals become available.



Proposed model of the dystrophin-glycoprotein complex. Dystrophin dimers link the cytoskeleton with the membrane glycoprotein complex by binding to actin at the amino terminus (N) and to the 50-, 43-, and 35-kDa transmembrane proteins at the carboxyl terminus (C).

From Ervasti, J.M., and Campbell, K.P. 1991. Cell *66:1121–1131. Copyright © 1991 by Cell Press.*

Molecular Studies of Calcium Channels and the Dystrophin-Glycoprotein Complex

Kevin P. Campbell, Ph.D.—Investigator

Dr. Campbell is also Professor of Physiology and Biophysics at the University of Iowa, Iowa City. He received his B.S. degree in physics from Manhattan College, his master's degree from the University of Rochester School of Medicine and Dentistry, and his Ph.D. degree from the Department of Radiation Biology and Biophysics at the University of Rochester. He did postdoctoral studies in the laboratory of David MacLennan at the Banting and Best Department of Medical Research, University of Toronto, before coming to Iowa.

CALCIUM ion (Ca^{2+}) functions as a ubiquitous intracellular messenger regulating a wide variety of cellular responses, including contraction, secretion, and cell proliferation. Two common pathways by which intracellular Ca^{2+} transients can be triggered have been identified. The first involves the influx of Ca^{2+} into the cytoplasm from the extracellular medium through specific plasma membrane channels, and the second involves the release of Ca^{2+} into the cytoplasm from intracellular stores through specific release channels.

Over the past 10 years, we have investigated the structure and function of the membrane components involved in Ca^{2+} fluxes across membranes. In particular, we have focused on identifying, purifying, and characterizing the membrane proteins that function as surface Ca^{2+} channels and intracellular Ca^{2+} release channels in excitable cells.

Neuronal Calcium Release Channels

A major area of research in my laboratory concerns the structure and function of intracellular Ca^{2+} release channels. In skeletal muscle, Ca^{2+} release from the sarcoplasmic reticulum initiates contraction. We previously purified a high-affinity ryanodine receptor from skeletal muscle sarcoplasmic reticulum and showed it to be identical to the sarcoplasmic reticulum Ca^{2+} release channel.

In neurons, inositol 1,4,5-trisphosphate (IP_3) is an important second messenger involved in Ca^{2+} release. However, physiological and pharmacological evidence has indicated the presence of non-IP_3–gated Ca^{2+} pools in neurons. Recently my laboratory also identified and purified a neuronal ryanodine receptor. The purified receptor was shown to be a homotetramer composed of protein subunits of approximately 500 kDa. Reconstitution of the purified ryanodine receptor into lipid bilayers has demonstrated that it functions as a caffeine- and ryanodine-sensitive Ca^{2+} release channel that is distinct from the brain IP_3 receptor. Immunoblotting experiments indicate that the brain receptor is more like the cardiac receptor than the skeletal receptor. Thus we believe that the brain ryanodine receptor may operate as a Ca^{2+}-induced release channel for intracellular Ca^{2+} pools in neurons. In the upcoming year further studies of this receptor should provide insights into its role in neuronal Ca^{2+} homeostasis and its distribution within the central nervous system.

Voltage-gated Calcium Channels

A second major area of research concerns voltage-gated Ca^{2+} channels in excitable cells. We have studied in skeletal muscle the dihydropyridine-sensitive Ca^{2+} channel, which is essential to excitation-contraction coupling. The skeletal muscle dihydropyridine receptor has been purified and consists of four subunits (α_1, α_2, β, and γ).

In neurons, voltage-gated Ca^{2+} channels exist as several types (L, N, T, and P) with different kinetic and pharmacological properties. Dihydropyridines bind specifically to L-type Ca^{2+} channels and alter their channel activity. For N-type channels, which are likely responsible for triggering neurotransmitter release at synapses, ω-conotoxin is largely specific. We are now using antibodies and cDNA probes to the various subunits of the dihydropyridine-sensitive channels to study ω-conotoxin–sensitive Ca^{2+} channels.

We have demonstrated that the brain ω-conotoxin–sensitive Ca^{2+} channel contains a component homologous to the β-subunit of the dihydropyridine-sensitive Ca^{2+} channel of skeletal muscle. We have also isolated a cDNA clone from a brain cDNA library encoding a protein with high homology to the β-subunit of the skeletal muscle dihydropyridine-sensitive Ca^{2+} channel. This brain β-subunit cDNA encodes numerous consensus phosphorylation sites, suggesting a role in Ca^{2+} channel regulation.

We are now purifying the ω-conotoxin–sensitive Ca^{2+} channel, using affinity chromatography, in order to analyze its subunit composition and to demonstrate that it is identical to the N-type Ca^{2+} channel. Experiments are also in progress to coexpress the brain β-subunit with the α_1-subunit

in order to study how the β-subunit regulates Ca^{2+} channel activity.

Dystrophin-Glycoprotein Complex

A third major project in my laboratory is aimed at understanding the function of dystrophin in normal muscle and determining how the absence of dystrophin leads to Duchenne muscular dystrophy (DMD). Dystrophin is localized to the inner surface of the sarcolemma in normal muscle but is absent in skeletal muscle of DMD patients and *mdx* mice. We previously showed that dystrophin is tightly linked to a large oligomeric complex of sarcolemmal glycoproteins. We have isolated a dystrophin-glycoprotein complex and have shown that it consists of cytoskeletal, transmembrane, and extracellular components. These data have allowed us to propose a model for the organization of the dystrophin-glycoprotein complex (see figure). The membrane organization of the dystrophin-glycoprotein complex and the high density of dystrophin in the sarcolemma membrane suggest that this complex could have an important structural role in skeletal muscle.

In the past year we have investigated the relative abundance of each of the components of the dystrophin-glycoprotein complex in skeletal muscle from normal and *mdx* mice. Our results demonstrate that all of the dystrophin-associated glycoproteins (DAGs) are significantly reduced in *mdx* skeletal muscle and suggest that the loss of DAGs is due to the absence of dystrophin and not to secondary effects of muscle fiber degradation. Furthermore, we recently showed that the absence of dystrophin in skeletal muscle from DMD patients leads to a dramatic loss of all the components of the dystrophin-glycoprotein complex. Thus the abnormal expression of the DAGs may play a crucial role in molecular pathogenesis in DMD.

In order to identify the normal function of the DAGs, we recently established, by cDNA cloning, the primary sequence of two components of the dystrophin-glycoprotein complex. The 43- and 156-kDa DAGs are encoded by the same mRNA, and post-translational modification of a 97-kDa precursor protein results in two mature proteins: the transmembrane 43-kDa DAG and the extracellular 156-kDa DAG. In addition, we have shown that the 156-kDa DAG binds laminin, a well-characterized component of the extracellular matrix. Our results demonstrate that the 43/156-kDa DAG (named dystroglycan) is a novel laminin-binding glycoprotein and suggest that the function of the dystrophin-glycoprotein complex is to provide a linkage between the subsarcolemma cytoskeleton and the extracellular matrix. Our findings strongly support the hypothesis that in DMD a dramatic reduction in a 156-kDa DAG leads to the loss of a linkage between the sarcolemma and extracellular matrix. This may render muscle fibers more susceptible to necrosis or may disrupt the integrity of muscle.

Our goal for the next year is to clone the other DAGs in order to express the entire complex in nonmuscle cells to study its function. We also plan to examine the possible involvement of the DAGs in other muscular dystrophies.

This work was also supported by the Muscular Dystrophy Association.

Gene Targeting

Mario R. Capecchi, Ph.D.—Investigator

Dr. Capecchi is also Professor of Human Genetics at the University of Utah School of Medicine and Professor of Biology at the University of Utah. He received his B.S. degree in chemistry and physics from Antioch College and his Ph.D. degree in biophysics from Harvard University, where he worked with James Watson. Dr. Capecchi remained at Harvard as a Junior Fellow of the Society of Fellows and then joined the Harvard faculty. Before moving to the University of Utah, he was Associate Professor of Biochemistry at Harvard School of Medicine. Dr. Capecchi is a member of the National Academy of Sciences.

GENE targeting—homologous recombination between DNA sequences in the chromosomes of mouse embyro-derived stem (ES) cells and newly introduced, exogenous DNA sequences—has been used to create mice with null (ablating) mutations in members of two families of genes. The first set of genes are members of the *int* proto-oncogene family (*int-1*, *int-2*, etc.). This set of genes is believed to be involved in localized developmental decisions mediated through cell-cell signaling. Their protein products resemble growth factors, and *int-2* is a member of the fibroblast growth factor family. These genes were initially identified through their pathological role in the genesis of mouse mammary carcinomas. We are addressing the question of their normal role during embryogenesis.

Targeted disruption of *int-1* (*wnt-1*) resulted in mice with a range of phenotypes from death at birth to survival to adulthood. Those mice surviving exhibited severe ataxia (loss of balance and coordinated movement). Those $int\text{-}1^{-}/int\text{-}1^{-}$ mice that died at or near birth showed severe abnormalities in the formation of the entire cerebellum and the major portion of the midbrain, whereas the defect in the survivors was restricted to the formation of the anterior region of the cerebellum. We have shown that a pre-existing mutant mouse, identified by its ataxic behavior, contains a frameshift mutation in the *int-1* (*wnt-1*) gene. This allele also appears to be a null mutation and exhibits the same range of variation in expressivity as our targeted null allele.

We have created mice with null mutations in *int-2*. These mice also exhibit a variable phenotype from death at birth to survival to adulthood. The survivors are even fertile. The $int\text{-}2^{-}/int\text{-}2^{-}$ mice have defects in the formation of the vestibule and cochlea, resulting in loss of balance and deafness. However, the degree of malformation of these inner ear compartments varies.

The second set of genes we are analyzing constitute part of the developmental program that specifies positional information along the anteroposterior axis of the early embryo. These genes, collectively known as the Hox genes, code for transcription factors in both the human and mouse. The 38 Hox members are present on four linkage groups of four separate chromosomes. The four linkage groups are believed to have arisen early in chordate evolution, as a result of quadruplication of a single ancestral group common to both vertebrates and invertebrates. Expansion of this gene complex may have had a critical role in the progression from invertebrates to vertebrates, by supplying the necessary complexity to this network of genes to accommodate the development of our complex body plan.

In *Drosophila* these genes act as master switches directing the course of morphogenic development of each parasegment. A mutation in one gene can result in dramatic homeotic transformations of one body part into another, such as the conversion of antennae to legs. Determining the function of the corresponding genes in mammals is just beginning.

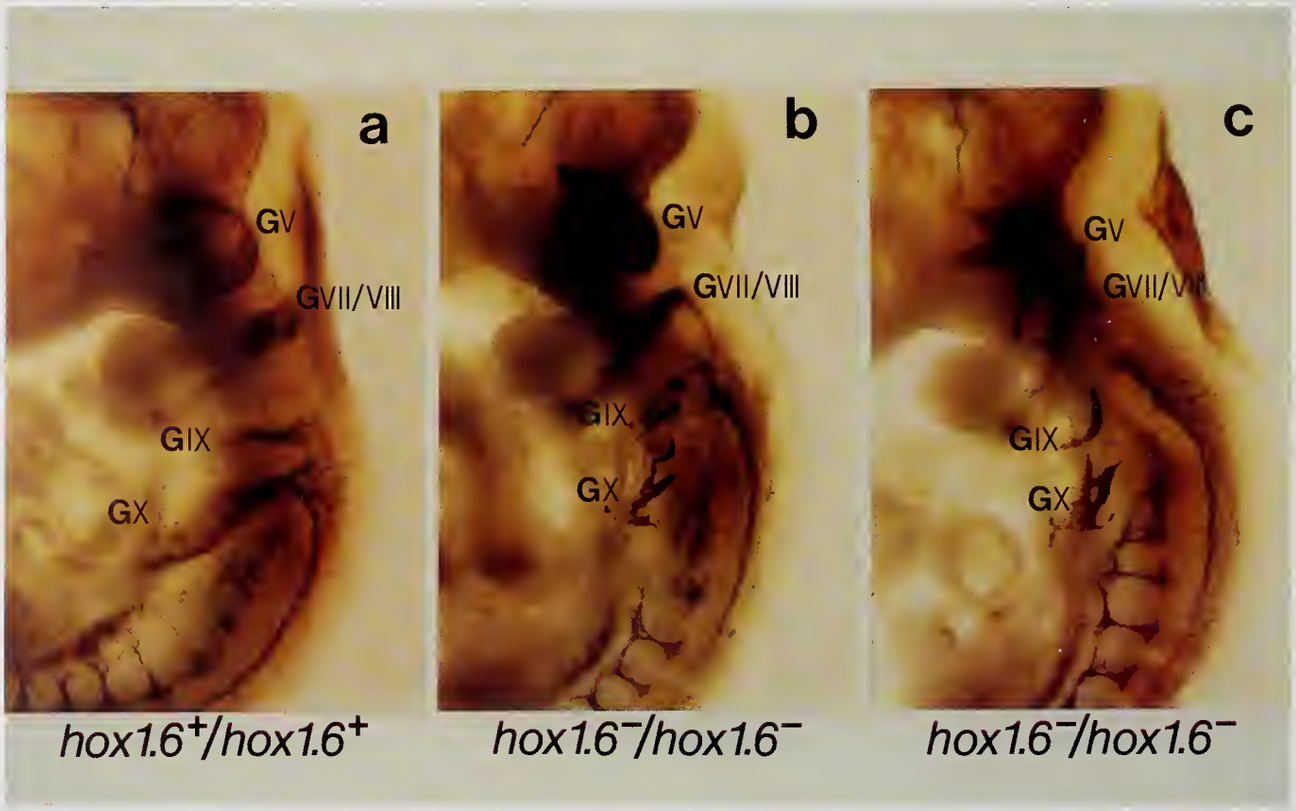
We have initiated a systematic genetic analysis of the Hox genes in mice. First, we are creating mice with null mutations in each of these genes to define their individual functions. From such an analysis, patterns should emerge that define the zones governed by these genes. Second, through epistasis and molecular analysis of combinations of *hox* mutations, we hope to define how this set of genes functions as a network to specify positional information along the body axis of the embryo. To date we have analyzed mice containing targeted disruptions in the two closely linked Hox genes, *hox-1.5* and *hox-1.6*.

Disruption of *hox-1.5* or *-1.6* resulted in mice with complex but regionally restricted developmental defects. Interestingly, the sets of defects associated with disrupting these two genes are distinct and non-overlapping. The $hox\text{-}1.5^{-}/hox\text{-}1.5^{-}$ mice are athymic, aparathyroid, and have reduced thyroid and submaxillary tissue. They also exhibit a wide spectrum of throat abnormalities (including shortened necks, abnormal larynx, truncated soft palate, and poor organization of throat musculature) and defects of the heart and major arteries. This collection of deficiencies is remarkably similar to those afflicting humans with the congenital DiGeorge syndrome.

On the other hand, $hox\text{-}1.6^{-}/hox\text{-}1.6^{-}$ mice

exhibit profound defects in the formation of the ear, hindbrain nuclei, and cranial nerves and ganglia. The glossopharyngeal and vagus nerves are poorly developed, and the preganglionic connections with the brain stem are not formed.

In both *box-1.5* and *-1.6* mutants, the affected tissues are formed by very different embryonic pathways, yet the two *box* genes are involved in executing each of those pathways. The nonoverlap of phenotypes associated with disrupting these two genes argues for their independent character. For example, if a nonredundant combinatorial code of *box* genes is used to specify each cell type in a given region, then the combined presence of both of these gene products does not specify any tissue.



Disruption of the hox-1.6 *gene results in regionally restricted developmental defects, including formation of the ear, cranial nerves, and ganglia, as well as the brain stem. This is illustrated by these embryos of E10.5 mice immunoreacted with a monoclonal antibody against the 155-kDa neurofilament protein. a) control* hox-1.6^{+}/hox-1.6^{+} *embryo; b, c) two mutant* hox-1.6^{-}/hox-1.6^{-} *embryos.*

Reprinted by permission from Chisaka, O., Musci, T.S., and Capecchi, M.R. 1992. Nature *355:516–520. Copyright © 1992 Macmillan Magazines Limited.*

Genetic Control of Pattern Formation in Drosophila

Sean B. Carroll, Ph.D.—Assistant Investigator

Dr. Carroll is also Associate Professor of Molecular Biology, Genetics, and Medical Genetics at the University of Wisconsin–Madison. He obtained his B.A. degree in biology from Washington University in St. Louis and his Ph.D. degree in immunology from Tufts University School of Medicine in Boston. He received postdoctoral training in developmental genetics working with Matthew Scott at the University of Colorado. In addition to his central work on pattern formation in Drosophila, *Dr. Carroll has also conducted basic research on new types of snake antivenoms, which are now under evaluation as potential pharmaceuticals. His honors include the NSF Presidential Young Investigator Award.*

INSECTS are the dominant group of animals on earth today. Nearly 1 million species have been classified among an estimated 20 million species extant. The innumerable sizes, shapes, and colors of insects have evolved from a basic segmented body plan consisting of three broad divisions—head, thorax, and abdomen—with three pairs of ambulatory legs and, in some orders, wings.

One insect, the fruit fly *Drosophila melanogaster*, has emerged as a key model for investigations into the genetics and molecular biology of animal development. An enormous amount of information has been gained about the organization of the *Drosophila* egg, the dynamic regulatory mechanisms guiding embryonic development, and the cellular and molecular processes involved in tissue differentiation and organogenesis. By addressing the intricacies of *Drosophila* development at the genetic, cellular, and molecular levels, one of the central puzzles of biology is being deciphered: How do complex animals form from a simple egg?

In addition, the stage is being set for comparative studies that will integrate the growing knowledge of *Drosophila* development with contemporary views of animal evolution. Efforts in our laboratory are aimed at both a detailed understanding of *Drosophila* embryology as a model for animal development and comparative studies of other insects (e.g., beetles and butterflies) as models for evolution.

Pattern Formation in *Drosophila* Embryos

The genetic control of pattern formation can be broken down conceptually into at least three phases. The first consists of a molecular prepattern, revealed as chemical changes that take place in different regions of the animal and foreshadow the cellular events to follow. For example, certain key regulatory proteins in *Drosophila* come to be expressed in stripes of cells encircling the embryo—stripes representing future segmental divisions. The second phase involves the specification of groups of precursor cells that populate the different tissues of the animal. For example, genes such as those of the *achaete-scute* complex (AS-C) are activated in clusters of cells that give rise to the central and peripheral nervous systems. Finally, in the third phase, these precursor cells divide and differentiate, giving rise to the full complement of specialized cells that make up different tissues and organs and express distinct structural genes to carry out their individual tasks.

From molecular prepattern to the formation of stem cells to the differentiation of their progeny, there is a flow of genetic information. The prepattern specifies the spatial domains of genes that are activated in stem cells, and these genes in turn regulate cell type–specific gene expression. Our laboratory is interested in the genetic basis of this information flow and focuses on three aspects of pattern formation.

First, we are studying the molecular regulation and function of the pair-rule genes, the first genes expressed in a segmentally repeating prepattern. Second, we are studying how the early prepattern genes regulate the expression of the stem cell–specific proneural genes of the AS-C. Third, we are trying to identify some of the molecular prepatterns that govern pattern formation in imaginal discs—the distinct pouches of cells that are set aside during embryogenesis, proliferate extensively during larval development, and give rise to adult structures such as the wing, leg, and eye.

In addition, we have recently initiated studies of other insects (guided by our knowledge of *Drosophila* genes) to compare the molecular aspects of segmentation in species that exhibit different styles of oogenesis and early development and to explore the genetic basis of wing development and evolution among the flying insects.

The Pair-Rule Segmentation Genes

During early *Drosophila* development, small batteries of genes are expressed in rapid succession to establish the segmental prepattern. The first genes expressed in a periodic pattern are the pair-rule genes, deployed in one stripe every two segments. The striped patterns of certain pair-

rule genes are generated by the aperiodic patterns of gap proteins, a handful of transcription factors expressed in broad, partially overlapping regions of the early embryo.

Three aspects of pair-rule gene regulation and function are under study. First, we wish to determine the molecular mechanisms of gap gene regulation of the pair-rule *hairy* gene. Second, we want to know how the pair-rule genes regulate the spatial expression of proneural genes during early formation of the nervous system. And third, we are investigating whether the same sort of segmentation genes function in insects with different styles of early development (e.g., beetles and butterflies). To this end, we are now isolating gap and pair-rule genes from other insects and arthropods to study their embryonic function.

Proneural Genes and Early Neurogenesis

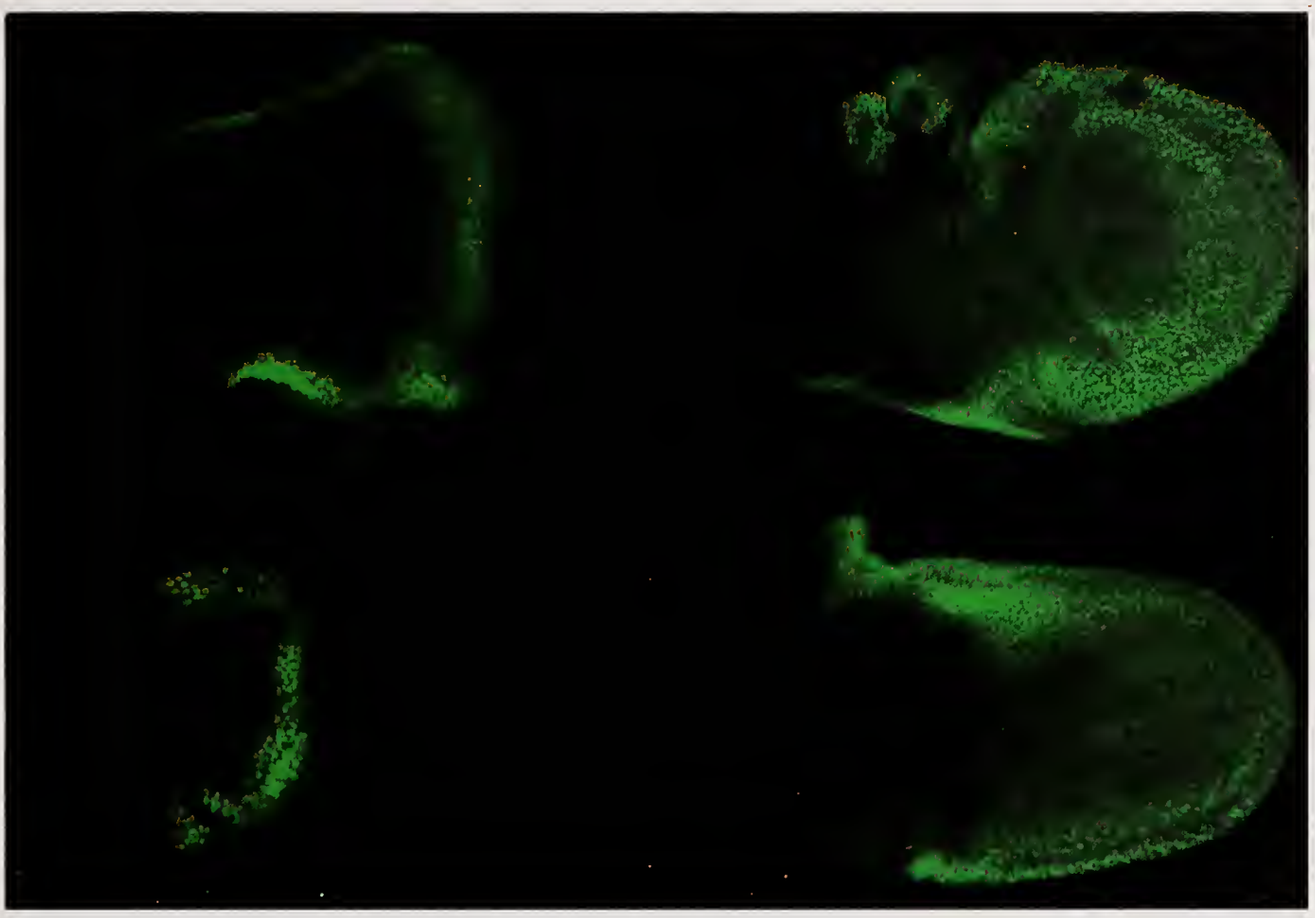
Once the embryo is subdivided by the segmentation genes, the formation of different tissues begins. One of the earliest events in embryogenesis is the segregation of the central nervous system precursor cells (neuroblasts) from the overlying ectoderm. The proneural genes, which are named for their role in promoting the neural over the epidermal pathway in the insect embryonic ectoderm, are expressed in segmentally repeating clusters of 4–6 cells. A single proneural-expressing neuroblast will segregate from each cluster.

We have recently determined that the establishment of proneural gene expression in these clusters is almost entirely regulated by up to eight pair-rule genes (all of which appear to be nuclear regulatory proteins) and an unknown number of genes acting along the dorsoventral axis. This demonstrates that the pair-rule genes are not simply involved in regulating the expression of other segmentation genes, but also direct the expression of genes that govern the behavior of small groups of cells. They serve to integrate the global regulatory system governing segmentation with the local regulatory system that specifies tissue architecture.

The pair-rule genes are not sufficient to specify individual cell fates, only the position of proneural cell clusters. Interactions between cells within each cluster determine their neural or epidermal fate. We have shown that the so-called neurogenic genes—a handful of loci that encode a variety of different proteins involved in cell surface interactions, signal transduction, and gene regulation—are required to single out just one neuroblast from the proneural cluster and prevent the other cluster cells from entering the neural pathway.

Development and Evolution of Insect Appendages

One of the best known but least understood aspects of *Drosophila* development concerns the morphogenesis of adult structures from the larval imaginal tissues. The eyes, wings, legs, antennae, and other structures are derived from sacs of cells that undergo extensive growth and morphogenesis during the larval and pupal stages. We are most interested in the development of the wing. We have recently shown that the product of the *vestigial* gene is precisely expressed in those cells of the imaginal wing disc that will form the actual flight appendage and is not expressed in the other disc cells that will form structural components of the thorax. In the absence of *vestigial* gene function, fruit flies are wingless, suggesting that *vestigial* plays a special role in the development of wings from imaginal tissues. Further studies into the regulation of *vestigial* expression in *Drosophila* may provide some interesting clues to the evolution and function of the *vestigial* gene in winged and wingless insects.



The nuclear vestigial *gene product controls wing and haltere development in the* Drosophila *embryo. Vestigial protein is specifically expressed in cells of the imaginal discs that will form the flight appendages—wing (top left) and haltere (bottom left). The nonexpressing cells of these discs will form structural parts of the thorax. During morphogenesis the* vestigial*-expression region everts (top right) and forms the wing blade (bottom right). In mutants lacking* vestigial *protein, no wing or halteres form.*

Research and photography by Jim Williams and Steve Paddock in the laboratory of Sean Carroll. From Williams, J., and Carroll, S.B. 1991. Genes Dev *5:2481–2495.*

Human Disease Gene Identification and Correction

C. Thomas Caskey, M.D.—Investigator

Dr. Caskey is also Professor of Molecular Genetics, Biochemistry, Medicine, and Cell Biology at Baylor College of Medicine. He received his M.D. degree at Duke University. His internship and residency training were in internal medicine, also at Duke; his postdoctoral training was at NIH, under the supervision of Marshall Nirenberg. Dr. Caskey is a past president of the American Society for Human Genetics and was named Distinguished Service Professor by the Board of Trustees of Baylor.

MOLECULAR genetics offers unprecedented opportunities for the discovery of disease genes, the development of simple DNA-based diagnostics, and correction of single-gene defects. Recent isolation of the gene responsible for the fragile X syndrome has led to improved diagnostic procedures and discovery of a novel mechanism for the occurrence of genetic disease. Further knowledge regarding the genetic mechanism and pathology of two other X chromosome disorders has also been gained. In addition, our laboratory has made significant progress in the development of gene replacement therapies for three diseases.

Disease Gene Identification and Diagnosis

The gene that contains the breakage point in fragile X syndrome has been isolated as the result of a collaboration with Stephen Warren (HHMI, Emory University) and Ben Oostra (Erasmus University). This gene contains a three-nucleotide sequence repeated in tandem, and the repeat region forms the fragile site itself. The number of repeats is variable and correlates with occurrence of the fragile X disorder. Normal individuals have approximately 6–50 repeats; affected individuals, over 200 repeats; and those with an intermediate number appear to harbor a "premutation" that is very likely to give rise to the full mutation by inheritance through females. This phenomenon of genetic disease caused by sequence amplification was previously unknown, but may prove to play a part in other inherited disorders.

Since the length of this particular region of DNA is an indicator of the affected or carrier status of the individual, DNA diagnosis of fragile X syndrome can be performed by Southern analysis, and more recently by the polymerase chain reaction. These DNA-based methods improve on cytogenetic diagnosis, particularly in the detection of female carriers and unaffected males who nevertheless can transmit the disorder.

Study of Genetic Disorders

The central region of the dystrophin gene is a "hot spot" for the deletion end-points causing Duchenne and Becker muscular dystrophies. This region was studied in greater detail to determine possible mechanisms that could explain why deletions were so frequent. In two independent patients with Duchenne muscular dystrophy, the ends of deletion lay within a transposable element (a portion of DNA that can move around the genome) in the dystrophin gene. It is likely that such transposable elements are involved in other deletions in this and other genetic disorders.

Lesch-Nyhan syndrome in humans results from lack of the enzyme hypoxanthine guanine phosphoribosyltransferase (HPRT). The neurological aspects of this incurable disorder are poorly understood. Study of the syndrome and development of therapeutic measures would greatly benefit from the generation of an animal model. Transgenic mice have been engineered to lack the uricase enzyme (which is absent in humans), and these are being bred with mice lacking the HPRT enzyme. This strategy has potential for generating a mouse model of the human Lesch-Nyhan syndrome.

Genetic Correction of Inherited Disease

Human genes can now be cloned, placed into defective (safe) viral vectors, and transferred into other cultured cells, embryonic cells, and animals. These encouraging developments increase the likelihood of successful gene replacement therapy. Our laboratory is developing technology toward that objective for three heritable diseases. Each disease offers different technical and strategic challenges.

Adenosine deaminase (ADA) deficiency is an inherited autosomal recessive disease. Bone marrow transplantation is curative, but carries the risk of a graft rejection. Administration of PEG ADA (the enzyme ADA attached to polyethylene glycol) on a continuing basis has provided improvement in the immunologic function of patients, but is not a cure.

The goal of this laboratory is the development of a clinical gene therapy protocol for ADA deficiency. This will require the treatment of human

cells with supernatant containing vector viruses (and no cells), rather than with virus-producing cells, because the former method is considered to be safer. Supernatant viral infection of bone marrow cells resulting in the acquisition of ADA activity can be achieved, but high efficiency requires the presence of other supporting cells, which is impractical in the context of an autologous bone marrow transplant. Attempts are under way to purify bone marrow stem cells, which are responsible for the regeneration of the entire hemopoietic system, since targeting this group of cells would provide a more efficient approach to the permanent repopulation of bone marrow with genetically corrected cells.

Ornithine transcarbamylase (OTC) deficiency, the most common urea cycle defect in humans, is inherited in an X-linked recessive manner. Coma, seizures, and retardation are the result of hyperammonemia secondary to the enzyme deficiency. Liver transplantation provides a once-per-lifetime cure, although again with the potential complications of a graft-versus-host reaction. This condition is poorly managed through dietary protein restriction and medical therapy, making the disorder another excellent candidate for gene therapy.

Two mouse models of the human disorder are available for developing gene therapy. In this laboratory, the OTC deficiency in one of the strains of mice has been corrected by retroviral delivery of the human OTC gene to the small intestine. In current studies toward the development of human gene therapy for OTC deficiency, retroviruses and defective (safe) adenovirus vectors are used to transfer the human gene to cultured liver cells.

Duchenne muscular dystrophy (DMD) is a severe disorder, also inherited in an X-linked recessive manner. Deficiency of the protein dystrophin leads to multiple muscle abnormalities and eventually death. The gene is large and complex. Several mouse models of DMD are available for study, and the dystrophin deficiency in one of these strains has been corrected by introduction of a construct expressing mouse dystrophin.

In collaboration with Helen Blau (Stanford University), myoblasts (precursor muscle fiber cells) are being isolated from very young patients with DMD. Different constructs will be tested for the transfer of the dystrophin gene and for creating gene-corrected myoblasts that can be returned to the patient. The gene is so large that currently used viral vectors will not accommodate it. Approaches being used to overcome this problem include modifying the vectors, truncating the gene, and using physical methods (that do not present a size limitation) for transferring the gene, such as electroporation or ballistic gene transfer (using the Du Pont "gene gun").

Enzymatic RNA Molecules and the Structure of Chromosome Ends

Thomas R. Cech, Ph.D.—Investigator

Dr. Cech is also American Cancer Society Professor at the University of Colorado at Boulder and Professor of Biochemistry, Biophysics, and Genetics at the University of Colorado Health Sciences Center, Denver. He received his B.A. degree in chemistry from Grinnell College and his Ph.D. degree in chemistry from the University of California, Berkeley. His postdoctoral work in biology was conducted in the laboratory of Mary Lou Pardue at the Massachusetts Institute of Technology. Dr. Cech is a member of the National Academy of Sciences. Among his many honors are the Lasker Award and the 1989 Nobel Prize in chemistry.

THE nucleic acids, DNA and RNA, are best known as the information-carrying molecules of a living cell. For example, a molecule of DNA or RNA might carry the instructions to build myosin, a protein involved in muscle movement, or pepsin, a protein enzyme that helps digest food. Our laboratory is investigating the non-informational roles of nucleic acids, situations in which a nucleic acid molecule has an important cellular function other than encoding a protein. In the area of RNA catalysis, we wish to understand how a folded RNA structure can have enzymatic activity. In the area of chromosome function, we are characterizing the DNA and associated protein necessary for proper maintenance of a chromosome end. In both projects, chemical and biological approaches are combined for fuller analysis of a biochemical problem.

RNA Catalysis

A cell must orchestrate thousands of chemical reactions in order to live, grow, and respond to its environment. These chemical reactions, rarely spontaneous, are usually catalyzed by macromolecules called enzymes. It was long thought that all enzymes were proteins. More recently we and others have found that a nucleic acid, RNA, can in some cases act as an enzyme.

The finding of RNA catalysis has several implications. First, it means that RNA is not restricted to being a passive carrier of genetic information but can participate actively in directing cellular biochemistry. In particular, many RNA-processing reactions are at least partly catalyzed by RNA. Second, the study of RNA enzymes, called ribozymes, may reveal hitherto unknown mechanisms of biologic catalysis. Third, ribozymes can be used as sequence-specific RNA cleavage agents *in vitro*, providing useful tools for study of RNA biochemically. Finally, on a more speculative note, RNA catalysis has the potential of providing new therapeutic agents. For example, ribozymes efficiently cleave and thereby destroy viral RNAs under controlled laboratory conditions, suggesting that they might be able to inactivate viruses in a living organism.

Many of our studies of RNA catalysis concern the *Tetrahymena* ribozyme, named for the single-celled animal from which it was originally isolated. This RNA enzyme is capable of cleaving other RNA molecules (substrates) in a sequence-specific manner. One of our objectives is to understand the mechanisms by which this RNA molecule acts as a catalyst. A second goal, in the area of structural biology, is to obtain a detailed picture of the active site of this ribozyme.

Last year we demonstrated that this ribozyme uses a novel mode of RNA recognition to bind its RNA substrate. In addition to the well-established mode of binding by formation of base pairs (as in the "ladder" of the DNA double helix), the ribozyme also binds two of the sugar groups that form the "backbone" of the RNA substrate chain. We have now located the other partner in this interaction, a specific adenine base within the RNA's catalytic core. In addition to support from HHMI, a grant from the National Institutes of Health supported a graduate student working on this project.

In a separate study, nucleic acid chemistry was used to change individual oxygen atoms to sulfur atoms near the cleavage site in the RNA substrate. With one of the sulfur-substituted RNAs, we observed a change in metal-ion specificity in the cleavage reaction. Thus we believe we have located one of the long-postulated active-site metal ions and have established that this RNA, like some protein enzymes, is a metalloenzyme.

In the structural area, we have developed a new technique for identifying parts of the folded RNA molecule that are in proximity to some known site. We attach a controllable agent of destruction to a small molecule known to dock in the ribozyme's active site, allow it to bind, initiate the chemical reaction, and map the sites of damage. The method has revealed convincing information about the higher-order folding of this RNA catalyst.

Telomere Structure

Unlike the circular chromosomes of bacteria, the chromosomes found in the nuclei of higher

organisms are linear DNA molecules. The ends of linear chromosomes, called telomeres, must be protected from degradation, and special features are required to ensure their replication. We are studying telomere structure and function, with special emphasis on the protein that caps off the ends of each chromosome.

Most cells have only a few dozen chromosomes and therefore not many telomeres. We chose to work with a ciliated protozoan, *Oxytricha nova*, because it has 26 million miniature chromosomes per cell. This facilitated our earlier purification of the telomeric protein and our cloning and sequencing of the genes that encode it. Similarities in DNA sequences in the telomeres of *Oxytricha* and those of higher cells, including human cells, give us reason to believe that our findings in *Oxytricha* will be of some generality.

Last year we used genetic engineering methods to produce the two subunits of the telomere-binding protein in bacteria. This has now allowed us to assess the individual contributions of the two subunits to the protein-DNA complex (work supported mainly by a grant from the National Institutes of Health). It has also allowed the purification of large amounts of the protein, enabling attempts to crystallize this complex, in collaboration with Steve Schultz's research group (University of Colorado). If successful, crystallization could lead to structure determination via x-ray diffraction, providing an atomic-resolution picture of the end of a chromosome.

Molecular and Cellular Physiology of Acute Inflammatory Cytokines

David D. Chaplin, M.D., Ph.D.—Associate Investigator

Dr. Chaplin is also Associate Professor of Medicine, Genetics, and Molecular Microbiology at Washington University School of Medicine and Assistant Physician at Barnes Hospital, St. Louis. He received his A.B. degree in biochemistry from Harvard University and his M.D. and Ph.D. degrees in cellular and developmental biology from Washington University. Following a residency in internal medicine at Parkland Memorial Hospital, Dallas, he received postdoctoral training in genetics at Harvard Medical School with Jonathan Seidman. He then returned to Washington University as a faculty member.

PROPER regulation of immune defenses requires the coordinate action of an array of different host responses. Many of these responses depend for their regulation on the interactions of several different types of immune cells. Some of these interactions are mediated by physical contact between the participating cells. Others, however, are mediated by soluble proteins that are released from one cell to act by binding to specific receptors on other cells. These soluble mediators have been designated cytokines.

Three cytokines appear particularly important for the initiation of acute inflammatory and immune responses: interleukin-1 (IL-1), tumor necrosis factor (TNF), and interleukin-6 (IL-6). IL-1, TNF, and IL-6 are pleiotropic cytokines. They are produced by a wide variety of cells, and they act on many tissues, showing a broad array of activities. In general, these activities are proinflammatory, favoring first the recruitment of white blood cells to sites of inflammation, then activation of these cells to express their full effector potential. This recruitment and activation is absolutely necessary for host defense against invading microbes; however, when it occurs inappropriately, it can lead to acute and chronic damage to self tissues.

Although a great deal of information has been accumulated concerning the activities of these cytokines when added to cultured cells or when injected into live animals, little is known about their true physiological function in normal host responses. In the case of IL-1, this uncertainty is underscored by studies of the cytokine's structure. First, biochemical and molecular genetic analysis of IL-1 has established that there are actually two forms of IL-1 (IL-1α and IL-1β), each the product of a discrete gene. The relative production of each form is variable, with some types of cells producing primarily IL-1α and some primarily IL-1β. Both forms are synthesized in the IL-1–producing cell as a larger precursor molecule (pro-IL-1α and pro-IL-1β), and both are cleaved to a smaller mature form that is found outside the cell. *In vitro*, mature IL-1α and -1β show apparently identical activities. Surprisingly, the two cytokines share only 25 percent amino acid sequence identity. Early studies from our laboratory showed that the genes encoding the two IL-1 proteins were generated by duplication of an ancestral IL-1–like gene, with subsequent extensive mutation to produce the existing IL-1α and -1β genes. We interpret the separate lineages of the two genes to indicate that there has been selective pressure for the independent evolution of each gene. Although this pressure probably acts at the level of unique functions for each of the two IL-1 molecules, *in vitro* studies have not revealed any functional differences between the two cytokines. This indicates to us that existing *in vitro* models do not adequately represent the normal physiology of IL-1 and that new models must be explored to disclose the functional differences that must exist *in vivo*.

Just as there are fundamental questions concerning the unique functions of each of the IL-1 proteins, there are basic questions concerning the mechanism by which these proteins exit the cell to perform their extracellular functions. Normally proteins that are destined to act outside the cell contain a characteristic signal peptide, a short amino acid sequence that directs them into the cellular secretory pathway. Neither IL-1 precursor protein contains a secretion signal peptide. Additionally, unlike conventionally secreted proteins, pro-IL-1α and -1β reside in the cellular cytoplasm, not associated with any membranous cellular compartment. Consequently it has been suggested that the IL-1 proteins are released via a novel secretory pathway.

Recent studies from our laboratory have yielded new insights into the mechanism of IL-1 release and maturation. In these studies we have shown that release of IL-1α and -1β from activated cells does not involve a unique secretory mechanism but occurs as a consequence of injury to the IL-1–producing cell. This conclusion is based on the observation that IL-1 release is accompanied by parallel release of other cytoplasmic proteins, such as the enzyme lactate dehydrogenase (LDH). Agents that enhance IL-1 release either induce increased IL-1 synthesis or increase

the degree of cellular injury and cause greater release of LDH. Conversely, agents that reduce IL-1 release either reduce IL-1 synthesis or enhance cell viability.

Although the release of IL-1 from IL-1–producing cells is not via a specific release pathway, the outcome of release can be modulated. This is demonstrated by studies in which IL-1–producing cells are deliberately injured. In the most general terms, there are two types of cellular injury. One type occurs when vital cellular functions are interrupted, leading to cellular necrosis. This occurs as a consequence of many stress injuries, including hypoxia, alterations in local pH, extremes of temperature, and physical attack by infectious lytic viruses. Morphologically, necrotic cells exhibit disruptions in their plasma membrane and disintegration of the intracellular organelles. The consequence of necrosis is the release into the extracellular space of all of the soluble cellular components. As expected, when an IL-1–producing cell sustains a necrotic injury, the IL-1 precursors are released along with all of the other cellular components; however, there is failure of processing of the IL-1β precursor to its mature form. This failure is important, because maturation is required for the IL-1β precursor to acquire receptor-binding activity (and, consequently, for the acquisition of all of the known IL-1 functional activities).

In contrast to cellular necrosis, cell death can occur in a regulated fashion. Normal development and tissue remodeling frequently require removal of otherwise viable cells. This occurs by activating the pathway of programmed cell death leading to apoptosis. Morphologically, apoptosis is very different from necrosis, with apoptotic cells showing retention of the integrity of their plasma membranes and maintenance of the structure of intracellular organelles. Portions of the cell are released as membrane-coated blebs and are subsequently ingested by phagocytic cells and digested. Apoptosis permits the termination of the life of a cell with minimal release of cellular debris into the extracellular environment. This pathway is activated during many stages of embryological development and to initiate normal cell turnover. In the immune system, removal of self-reactive T lymphocytes in the thymus occurs by apoptosis, as does the death of target cells attacked by cytotoxic lymphocytes. When this pathway of programmed cell death is activated in macrophages expressing the IL-1 precursors, there is striking, rapid, and complete maturation and release of both forms of IL-1. Thus, in cells expressing only IL-1β (such as blood monocytes and certain cells in the spleen and thymus), apoptosis—in contrast to necrosis—affords an efficient method for generating bioactive IL-1.

These results have given us a new appreciation of IL-1 as a physiological signal of cell and tissue injury. IL-1 does not appear to be released from the cell that produces it unless a significant cell injury occurs. If the injury is necrotic in nature, then IL-1β is released in only its inactive precursor form. If, on the other hand, the injury leads to apoptosis, then mature bioactive IL-1β is released in a form ready to activate the systemic responses that lead to tissue repair.

Hormonal Regulation of Gene Expression

William W. Chin, M.D.—Investigator

Dr. Chin is also Associate Professor of Medicine at Harvard Medical School and Senior Physician at Brigham and Women's Hospital, Boston. He obtained his undergraduate degree in chemistry from Columbia College and his M.D. degree from Harvard Medical School. His postdoctoral work was performed with Jacob Maizel and Philip Leder at NIH and with Joel Habener at Massachusetts General Hospital, Boston. His awards include the Bowditch Lectureship Award of the American Physiological Association, the Van Meter-USV Award of the American Thyroid Association, and the Outstanding Investigator's Award from the American Federation for Clinical Research.

HORMONES, key players in the endocrine and nervous systems, are produced by specialized tissues in the body, effect extracellular communication, and regulate cellular function. Our recent studies have considered the molecular mechanisms of gene regulation by a specific hormone, thyroid hormone.

Thyroid Hormone Action

Our early work has focused on the regulation of the thyrotropin (thyroid-stimulating hormone [TSH]) subunit genes by thyroid hormones. TSH is a polypeptide hormone that is produced and secreted by a single cell type in the anterior pituitary gland. It comprises two different sugar-containing polypeptide subunits, α and β, which are encoded by genes located on different chromosomes. Importantly, TSH stimulates the thyroid gland to produce the thyroid hormones, T_3 and T_4. These modified amino acids regulate metabolism and gene expression in almost every cell of the body. To maintain a constant level of T_3 and T_4 in the bloodstream, these hormones act back on the pituitary to decrease TSH production, including subunit gene expression, and secretion. Hence TSH and thyroid hormones are involved in a classic negative-feedback regulatory system.

To understand further the molecular mechanisms involved in the negative regulation of TSH synthesis by thyroid hormones at the transcription level, we have isolated and analyzed the genes encoding the α- and β-subunits of TSH in the rat and identified putative thyroid hormone–responsive elements (TREs). These TREs, which mediate a negative regulation by the thyroid hormones and thyroid hormone receptors (TRs), contain several consensus TR-binding half-sites in a direct repeat orientation and are located immediately upstream of the TATA box (α-subunit) or near the start of transcription (TSHβ subunit).

Multiple Thyroid Hormone Receptors and Other Nuclear Factors

Thyroid hormones, in general, act at the cellular level by entering the cell either as T_3 or T_4. T_4 may be converted to T_3 in certain tissues, such as the pituitary. T_3 then enters the nucleus, where it interacts with the TR, a protein encoded by the proto-oncogene c-*erbA*. The T_3-TR complex then binds directly to cis-DNA elements within thyroid hormone–responsive genes to activate the appropriate responses.

The TRs are encoded by two genes, α and β, each expressing at least two related molecules obtained by alternative promoter choice and splicing. The rat TRα gene encodes a bona fide TR, TRα1, and a related but non-T_3-binding form, c-*erbA*α*2*. The rat TRβ gene encodes two functional TRs, TRβ1 and TRβ2, which are identical except for different regions amino terminal to the DNA-binding domain. Remarkably, TRβ2 was most highly expressed in the pituitary gland. This observation has potential physiological relevance because of the major role of thyroid hormones in regulating important hormone genes in the pituitary, including TSH. The finding also stands in contrast to the general tissue distribution of the other receptor forms.

We recently demonstrated the existence of another nuclear protein(s) that can augment the binding of TRs to various TREs. This ubiquitous 60- to 65-kDa protein (TRAP; TR auxiliary protein) interacts with TR to form a heterodimer and also binds specific sequences within the TRE. Several regions of the carboxyl-terminal or ligand-binding domain of the TR are important for this heterodimerization. Preliminary results show that the TR-TRAP interaction may be critical in thyroid hormone–mediated transactivation. Furthermore, the recent identification of the retinoid X receptor as a potential TRAP by several groups fuels additional excitement in this area.

In summary, the thyroid hormone receptor family is complex. At least three biologically active forms are expressed in a tissue-specific fashion, and another form may have an important role in modulating the effects of the others. Knowledge of the functions of these receptor isoforms and their interactions with other nuclear factors will be critical for our full understanding of thyroid hormone action.

Thyroid Hormone Action at the Post-transcriptional Level

In addition to the well-known transcriptional effect of thyroid hormone on TSH synthesis, we have recently examined the effect of thyroid hormone on the half-lives of TSH subunit mRNAs *in vitro* and *in vivo*. The turnover rate of the TSHβ mRNA was greater in the presence of thyroid hormone. This phenomenon was associated with a shortening of the poly(A) tail of the mRNA. These events occurred in a dose- and time-dependent fashion. Thus thyroid hormone reduces the half-life of the TSHβ mRNA and causes a concomitant decrease in its poly(A) size. These data indicate that thyroid hormones may also exert their effects on gene expression at the post-transcriptional level.

Summary

Our work has focused on the molecular mechanisms of thyroid hormone action, with a focus on the regulation of TSH gene expression by thyroid hormones, including the cis-elements and trans-factors that are involved in these processes. The effects of these hormones are manifest at both the transcriptional and post-transcriptional levels. We hope that our efforts will provide insight into the hormonal regulation of gene expression and cellular function in normal and pathological endocrine states and in cancer.

Technologies for Genome-sequencing Projects

George M. Church, Ph.D.—Assistant Investigator

Dr. Church is also Assistant Professor of Genetics at Harvard Medical School. He received his B.A. degree in zoology and chemistry from Duke University and his Ph.D. degree in biochemistry and molecular biology from Harvard University. Before moving to Harvard Medical School, Dr. Church was a scientist at Biogen Research Corporation and a Life Sciences Research Foundation Fellow in the Department of Anatomy at the University of California, San Francisco.

THE study of the linear sequence of bases in genomic DNA and messenger RNA is steadily gaining recognition, in part as a result of the increasing ease and advantage of using shared computer databases to find connections among distant concepts and distant biological systems. For example, connections have been found between human oncogenes and yeast transcription factors, between differentiation antigens and bacterial chaperone proteins, and between developmental regulatory genes and bacterial DNA-binding proteins.

Unfortunately, these database searches are frequently unsuccessful, because not all classes of genetic elements are represented. The complete sequence of a few small genomes should rectify this. Sequencing projects have begun for genomes of various bacteria (*Mycoplasma*, *Mycobacteria*, *Escherichia*, and *Thermococcus*), a yeast (*Saccharomyces*), a plant (*Arabidopsis*), and a worm (*Caenorhabditis*), chosen for their well-studied genetics, their small genome sizes, and their representation of all major branches of the evolutionary tree. The genome closest to completion is that of *Escherichia*, with about 38 percent of its 4.7 million base pairs already in the database, through the effort of 2,000 researchers.

To improve the accuracy and efficiency of these projects, we have developed new sequencing technologies. One, called multiplex sequencing, is a way of keeping a large set of DNA fragments as a precise mixture throughout most of the steps of sequencing. Because each mixture can be handled with the same effort as a single sample in previous methods, more fragments can be handled.

The mix is deciphered by strategically tagging the fragments at the beginning with unique bits of DNA and then, at the end, hybridizing to the sequencing reactions complementary bits of DNA that have been spread out by size and immobilized on large membranes. This method also improves the accuracy, since the mixtures contain internal standards of known sequence that help in the computer analysis of the film data.

The number of probings obtainable per membrane represents the increased efficiency factor of this method. This number exceeds 50 now (the higher the better) and is likely to increase. We have designed and tested devices to facilitate most of the steps in multiplex sequencing, including DNA preparation, sequencing reactions, gel loading, hybridization, film exposure, and film reading. All of these devices have been applied to collect over 1 million bases of raw data and are undergoing further development. Multiplexing has also allowed chemiluminescent detection to replace the radioactivity normally used in DNA sequencing, reducing exposure times 10-fold. To fill in specific gaps in the sequences, we have devised multiplex oligonucleotide synthesis for use in multiplex DNA sequence walking strategies.

Toward the goal of modeling cell structure and gene expression, we have searched for abundant cellular proteins that have nonetheless eluded the extensive biochemical and genetics studies of *Escherichia coli*. This has been done by systematically correlating amino-terminal protein sequence data obtained from two-dimensional gel spots with the DNA sequence and two-dimensional gel databases. Of 300 sequences determined so far, over 50 are candidates for such major novel proteins.

We have extended our methods for detecting *in vivo* molecular interactions by analyzing the protection of individual DNA bases from enzymatic methylation. DNA protein interactions involved in cAMP and pyrimidine feedback regulation have been studied in this way.

In the future, with new sequencing technologies such as automated multiplex sequencing, with examples of most of the basic genetic modules, and with an eye to sequence elements conserved among species, the analysis and modeling by investigators worldwide of human sequences and genetics should become more manageable.

Molecular Genetics of Limb Development in Drosophila

Stephen M. Cohen, Ph.D.—Assistant Investigator

Dr. Cohen is also Assistant Professor in the Department of Cell Biology and in the Institute for Molecular Genetics at Baylor College of Medicine. He obtained his Ph.D. degree at Princeton University, working with Malcolm Steinberg. He did postdoctoral work with Harvey Lodish at the Whitehead Institute and with Herbert Jäckle at the Max Planck Institute for Developmental Biology, Tübingen, and at the University of Munich.

MY laboratory is interested in the processes by which genetic information can be used to organize the three-dimensional pattern of the body during embryonic development. The particular problem on which we have focused involves the organization of the limbs in the fruit fly *Drosophila*. We would like to understand the genetic and molecular mechanisms that the fly embryo uses to determine where the limbs will develop with respect to the rest of the body pattern. We are investigating how the appropriate embryonic cells become committed to develop as limb precursors and subsequently how these cells cooperatively organize the spatial pattern of the leg.

Specification of Limb Cell Identity

Using classical genetics, we have identified the gene *Distal-less*, which seems to function as a critical genetic switch that initiates limb development in the embryo. Embryos that lack *Distal-less* gene function do not develop larval or adult limbs. These observations indicate that *Distal-less* activity is critically required but do not tell us what the gene is actually doing. Using molecular probes, we can visualize the RNA product of the *Distal-less* gene in the embryo and in the developing adult limbs. All of the cells that will go on to form limb structures express *Distal-less*. An additional genetic experiment indicates that these cells need the *Distal-less* gene product to be limb cells. If the gene is removed from a group (or clone) of cells that were already committed to develop into limb structures, these cells instead become body wall cells. *Distal-less* therefore specifies their identity as limb cells. *Distal-less* is a member of a family of important regulatory genes that encode a sequence-specific DNA-binding motif called the homeodomain. *Distal-less* therefore can be presumed to define cell identity by regulating the expression of other genes. We wish to identify the transcriptional targets through which *Distal-less* acts.

How Does the Embryo Position Its Limbs?

The critical first event in limb development is determining the position at which the legs will develop. We are using a combination of genetic and molecular methods to identify the source of the instructions specifying the identity of the leg cells and the nature of the response that this signal elicits. In the *Drosophila* embryo, the legs begin as clusters of cells. These clusters are established at a well-defined location within each thoracic segment, with respect to both anterior-posterior and dorsal-ventral position. One way to mark a unique point is to draw two intersecting lines. This rather simple solution appears to describe what the embryo does to identify the cell clusters. Thus the spatial cues that the embryo uses to position its legs appear to be stripes of cells that act as sources of secreted intercellular signals. The earliest detectable consequence of this signaling process is that presumptive leg cells express the *Distal-less* gene.

Our previous work identified the segmentally repeated stripes of cells expressing the *wingless* gene as one of the signaling centers responsible for initiating leg development. Stripes of cells expressing *wingless* bisect the nascent leg primordia. The *wingless* gene encodes a secreted intercellular signaling molecule related to the vertebrate oncogene INT-1. The secreted *wingless* protein is required to signal nearby cells to express *Distal-less* but is not sufficient to do so. An intersecting stripe of information is required to specify the cluster. The *decapentaplegic* gene is a good candidate to encode this second signal. A stripe of cells expressing *decapentaplegic* intersects the *wingless* stripe at precisely the position where the leg is formed. The *decapentaplegic* gene encodes a product homologous to the transforming growth factor-β (TGF-β) class of secreted signaling molecules. A group of cells that lies near the intersection of these two stripes should receive both signals. Although we are not yet certain whether *decapentaplegic* plays a direct role in this process, the simultaneous receipt of the *wingless* and TGF-β signals should, in principle, be sufficient to specify the identity of the leg precursor cells.

Why Do Insects Have Six Legs?

Insects develop legs only in their thoracic segments, while more primitive arthropods may develop abdominal legs as well. A class of genes known collectively as homeotic genes distinguish segments from one another. The homeotic genes of the Bithorax complex act as negative regulators of thoracic segment identity. The Bithorax genes repress *Distal-less* expression and block leg development in abdominal segments. According to our model, the intercellular signal resulting from the intersection of *wingless*- and *decapentaplegic*-expressing cells should in principle be capable of eliciting limb development in the abdomen. The signaling sources are expressed in precisely the same pattern in thoracic and abdominal segments. We have found that this response is blocked by the Bithorax genes. Removing Bithorax genes permits *Distal-less* expression in the abdominal segments and leads to ectopic limb development in the abdomen. Therefore, the signal to specify leg primordia in the abdominal segments exists, but the consequences of signaling are overridden.

Molecular analysis of the control elements that regulate *Distal-less* expression in the embryo provides compelling evidence that the intercellular signal is actually received and correctly interpreted by cells in the abdominal segments, just as in the thorax. We have identified a small piece of DNA, known as an enhancer, that acts as the target of this signal. By functional dissection of this enhancer we have identified a regulatory element that mediates the repression of limb development in the abdominal segments. Removing a small piece of DNA renders the enhancer element insensitive to the repressor genes of the Bithorax complex but has no effect on its response to the activation signal. Gene expression driven by this mutant version of the enhancer element is therefore constitutively "de-repressed" in the abdomen. These observations demonstrate that the *wingless*- and *decapentaplegic*-dependent activation signal is both received and correctly interpreted by the would-be abdominal leg cells. Repression of leg development must therefore lie in the prevention of *Distal-less* expression by these cells. These observations strongly suggest that turning on the *Distal-less* gene in these cells is the only signal necessary to specify their identity as leg cells.

Proximal-Distal Pattern Formation in the Leg

As mentioned above, flies that lack *Distal-less* gene activity do not develop any leg structures. Flies in which *Distal-less* gene activity is impaired, but not eliminated, develop abnormal legs. *Distal-less* mutant legs are foreshortened along the proximal-distal axis. The characteristics of different *Distal-less* mutations tell us that the amount of activity of the gene is important in controlling the range of structures that the leg can develop. Distal parts of the leg require more *Distal-less* gene activity than do proximal parts to develop normally. These observations suggest that *Distal-less* may play an important role in organizing the proximal-distal axis of the leg.

Early in development, presumptive leg cells express *Distal-less*, whereas presumptive body wall cells do not. As the leg matures, this simple pattern transforms into a graded distribution of *Distal-less* RNA across the developing limb. The distal-most region of the limb expresses a high level of the gene product, intermediate levels express lower levels of *Distal-less* RNA, and proximal regions express little or no *Distal-less* RNA. These observations are particularly intriguing in view of the regional differences in the requirement for the activity of the gene along the leg. We are interested in understanding the transition from the initial simple, two-state system of the embryo to the complex, graded distribution of the gene product that we see at later stages.

Understanding pattern formation in the developing fly leg will teach us about fundamental mechanisms important for the development of vertebrate embryos. We are studying a system in which intercellular signaling molecules are used to assign cells to functional developmental units. Although the particular details of the systems vary, the fundamental principles and molecular mechanisms will prove to be of wider applicability. In this context we are intrigued by the role that *Distal-less* plays as a pattern organizer at the genetic level and its prospective role as a regulator of gene expression at the molecular level.

Tracking Genes That Cause Human Disease

Francis S. Collins, M.D., Ph.D.—Investigator

Dr. Collins is also Professor of Internal Medicine and Human Genetics at the University of Michigan Medical School. He received his Ph.D. degree in physical chemistry from Yale University and his M.D. degree from the University of North Carolina. After completing his internship and residency in internal medicine at the North Carolina Memorial Hospital, he went on to a fellowship in human genetics at Yale. He recently received the Young Investigator Award of the American Federation of Clinical Research and the National Medical Research Award of the National Health Council and was elected to the Institute of Medicine and the American Association of Physicians.

THE theme of our laboratory is the study of human genetic disease at the molecular level. Our goal is to identify genes involved in specific genetic disorders, to define their structure and function, to understand the control of their expression, and to use this information to develop potential new therapies.

Over 4,000 genetic disorders are listed in the most recent edition of *Mendelian Inheritance in Man*. For the majority of these, the normal function of the gene involved is not known. The identification of disease-causing genes without knowledge of their protein product or its normal role, now referred to as positional cloning, is a major endeavor in our research.

Only recently has it become possible to identify such genes, and the process is still laborious. First the gene must be mapped to a specific human chromosome, using a process known as linkage analysis. This involves identification of families with the disorder and analysis of DNA from these families with a panel of probes from all parts of the human genome. The probes are used to establish the chromosomal location of a DNA sequence that may have been inherited in association with a disease gene that will probably lie close to it. Such sequences, or "markers," can pinpoint the chromosome on which the disease gene resides. Additional probes from that chromosome can then be tested to identify markers that are even closer. It is often possible to narrow the responsible region to about 1 percent of a particular chromosome. Although this is a major achievement, such a region may yet contain 30 to 50 genes, only one of which is responsible for the disease. Thus additional refinements must be made before candidates for the responsible gene can be identified.

In 1989, as part of a collaborative effort with investigators at the Hospital for Sick Children in Toronto, we were successful in identifying the gene for cystic fibrosis (CF), a common severe genetic disease characterized by lung infections, pancreatic insufficiency, and elevations in sweat chloride concentration.

Although one mutation (ΔF508) is responsible for about 75 percent of CF chromosomes in the United States, numerous other mutations have now been found in other individuals with the disease. This year we identified one mutation that gives rise to such a mild form of the disorder that it was not recognized as being caused by this same gene, because the affected individuals had normal sweat chloride. This observation indicates that mutations in the CF gene may have a broader clinical spectrum than previously suspected.

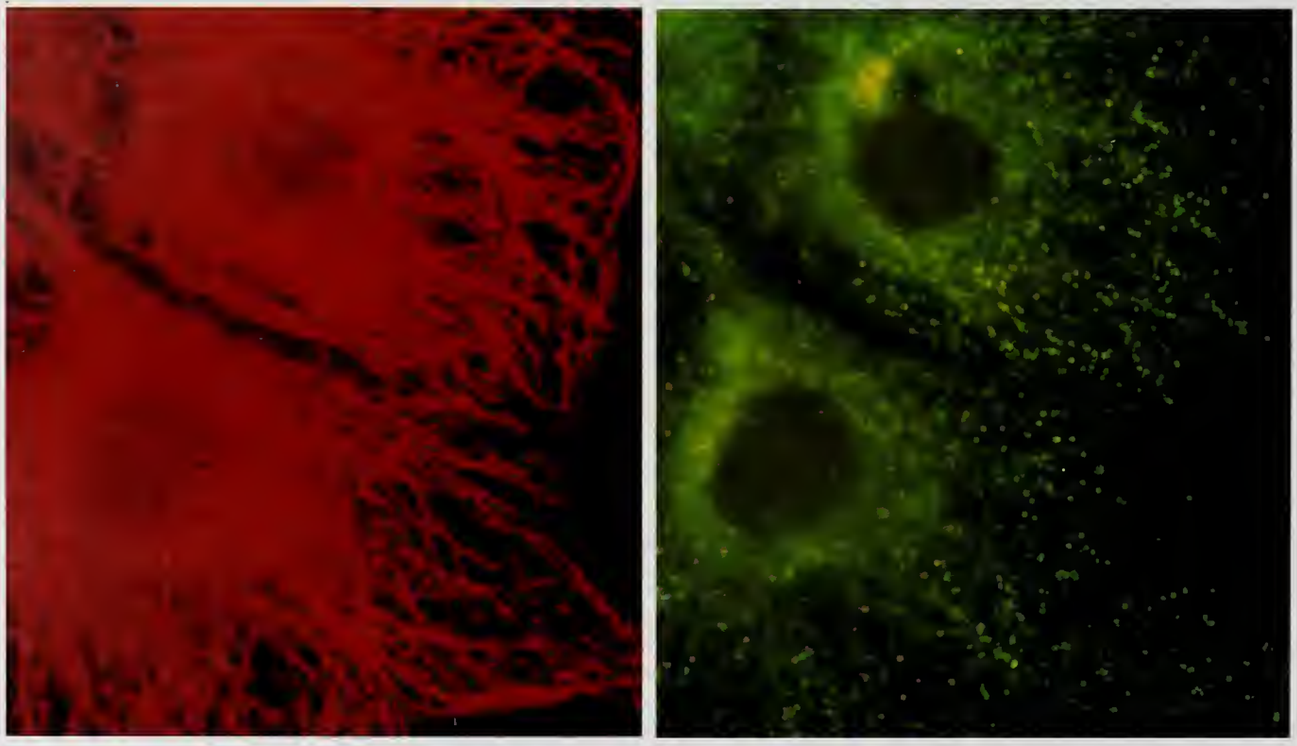
Another important observation in CF this year has been the demonstration by our group that the ΔF508 mutation does not completely inactivate the gene. In fact, using a frog oocyte system, we were able to show that the ΔF508 protein can be activated to almost normal levels by using drugs that raise levels of cAMP, an intracellular second messenger. This suggests that drug therapy aimed at this second messenger might benefit patients with the disease by activating their defective CF protein.

Another major project in our laboratory is an investigation of von Recklinghausen neurofibromatosis (NF1), a common genetic disorder sometimes incorrectly referred to as the Elephant Man disease. Having identified this gene in 1990, we are now intensively attempting to identify mutations that cause the disease and to determine the normal function of the protein product of this gene. Previous studies have demonstrated that this protein interacts with another class of proteins (encoded by the *ras* genes) involved in the regulation of cell growth. Recent data from our group have shown that the NF1 protein also interacts with the cytoskeleton and can be seen to colocalize with microtubules using immunofluorescence techniques. This unexpected observation suggests that the protein product of the NF1 gene may play a crucial role in the regulation of cell division. In the longer term, it is hoped that this improved understanding of the biology of the disease will lead to improved therapies.

The Huntington disease (HD) gene has been

localized by a collaborative group, which includes our laboratory. It is now known to lie in a region near the tip of the short arm of chromosome 4, within an interval of approximately 2 million base pairs. Using yeast artificial chromosome cloning, we have generated an overlapping set of genomic clones from this region and are utilizing a variety of novel technologies to search for candidate genes.

A new project in our laboratory this year is an intense effort to identify the gene responsible for early onset breast cancer, which has recently been mapped by Mary-Claire King and her colleagues (University of California, Berkeley) to chromosome 17. Using genetic linkage analysis in families, as well as the analysis of tumors, we are narrowing down the responsible interval. Significant portions of the candidate interval have now been cloned in yeast artificial chromosomes. Utilizing the relatively new technology of physical microdissection of chromosomes, we have been able to produce many new markers in this region of the long arm of chromosome 17; thus the opportunity to identify a major gene contributing to this common and devastating disease in Western society is now quite real.



Colocalization of neurofibromin (the normal protein product of the neurofibromatosis gene) and microtubules. The same rodent cell has been double-labeled and visualized by fluorescence microscopy. Red shows staining with an antibody directed against tubulin, the major component of cytoplasmic microtubules. Green shows staining of neurofibromin that appears to decorate the microtubules.

Research of Paula Gregory, David Gutmann, and Francis Collins.

Development of the Immune System

Max D. Cooper, M.D.—Investigator

Dr. Cooper is also Professor of Medicine, Pediatrics, and Microbiology at the University of Alabama at Birmingham. He received his M.D. degree and specialty training in pediatrics at Tulane Medical School and his postdoctoral training in immunology at the University of Minnesota.

INFORMATION obtained from studies of immune system development in a variety of vertebrate species is used to investigate diseases of the immune system in humans. We are particularly interested in the pathogenesis of immunodeficiency diseases and lymphoid malignancies.

Comparative studies in birds and mammals initially revealed the separate developmental pathways of T and B lymphocytes, the two major types of immunocompetent cells. T cells provide help for antibody-producing B cells and are primarily responsible for immunity against viruses and fungi. We wish to know whether the thymus is the only source of T cells in vertebrates or whether they may also arise in other tissues.

The central tissues may vary for production of B cells, the source of antibodies. In birds, immunoglobulin (Ig)-bearing B cells are derived from the hindgut bursa of Fabricius, whereas in mammals such cells are generated in blood-forming organs, mainly the fetal liver and adult bone marrow. Multipotent stem cells in these tissues use an elaborate gene program to generate millions of T and B lymphocyte clones, each expressing a T cell receptor (TCR) or Ig receptor of different antigen specificity. These are seeded via the bloodstream to the peripheral tissues, where they execute immune surveillance of foreign and self antigens.

Inherited or acquired gene defects may specifically alter growth or maturation of these cell lines to cause immune system dysfunction or malignancy.

Comparative Analysis of T Cell Development

We have embarked on a comparative analysis of T cell development in representative mammalian, avian, and amphibian species. Our studies in birds reveal remarkable evolutionary conservation of the pattern of T cell development found in mammals, including the sequential development of cells bearing either the $\gamma\delta$ or $\alpha\beta$ TCR.

Cells expressing the $\gamma\delta$ TCR are generated first during ontogeny. They may not undergo clonal selection during their intrathymic development. They migrate preferentially to splenic red pulp regions and the intestinal epithelium. Unlike $\alpha\beta$ T cells, the $\gamma\delta$ T cells cannot recognize class II molecules of the major histocompatibility (MHC) gene complex to initiate a graft-versus-host (GVH) attack. Analysis of their physiological role is facilitated by their relative abundance in birds (25–50 percent of T cells), where we are examining their development and functional capabilities.

In collaboration with Craig Thompson (HHMI, University of Michigan), we have used the TCR2 and TCR3 antibodies to identify two discrete sublineages of $\alpha\beta$ cells that use different variable-region gene families: Vβ1 and Vβ2. The avian Vβ1 and Vβ2 genes contain highly conserved sequences that distinguish the two major subgroups of mammalian Vβ genes. Birds thus provide a simple model system for study of the functional significance of these prototypic gene families.

In birds, as in mammals, TCRβ diversity is created largely by nucleotide sequence variations in the joints between rearranged Vβ, D (diversity), and J (joining) genes. Since Vβ1 and Vβ2 genes both combine with the same D and J genes, the V$\beta1^+$ and V$\beta2^+$ T cells should be capable of recognizing the same spectrum of antigenic peptides. However, the Vβ1 and Vβ2 genes have little sequence homology (less than 30 percent), so their protein products may interact differently with peptide-presenting MHC molecules.

Indeed, although both V$\beta1^+$ and V$\beta2^+$ T cells can cause GVH disease in recipients with the same MHC class II genes, they do so with different efficiency. Selective inhibition of Vβ1 T cell development revealed that only the V$\beta1^+$ T cells can help B cells produce protective IgA antibodies along the mucous surfaces of the body. Current studies explore in more detail the functional capabilities of the Vβ1 and Vβ2 subpopulations of avian $\alpha\beta$ T cells.

To analyze T cell development in an amphibian, we have made monoclonal antibodies that identify $\gamma\delta$ and $\alpha\beta$ TCR candidates expressed by separate lymphocyte subpopulations in *Xenopus laevis*. Frog T cells bearing the putative $\gamma\delta$ and $\alpha\beta$ TCR homologues exhibit the same tissue distri-

bution patterns shown previously for their avian and mammalian counterparts. Cloning of the frog TCR genes is one current focus, and the thymus dependence and functional capabilities of the two sublineages are being explored. The goals of these studies are to understand the evolutionary strategy for generating T cells that can discriminate between self and nonself and to gain fresh clues for some of the unresolved mysteries of the human immune system.

B Cell Development

Bone marrow stromal cells can influence the growth and development of B cell precursors, either by direct cell contact or via soluble products like interleukin-7 (IL-7). We are examining cell surface molecules through which pre–B cells may receive these environmental cues. One such candidate molecule is an ectoenzyme named aminopeptidase A (APA), the increased expression of which is induced by exposure to IL-7. We are testing the hypothesis that IL-7 is either a ligand or a catalytic substrate for APA. We are using a mouse gene probe to identify the human gene for APA for sequence analysis, with the goal of understanding its expression as a function of normal and leukemic pre–B cell differentiation.

Before pre–B cells rearrange their light-chain genes to become B cells, they express a complex receptor that is composed of surrogate light-chain proteins (V_{preB} and $\lambda 5$), μ heavy chains, and two small transmembrane proteins called α and β. We have produced monoclonal antibodies against exposed configurations of these protein receptor units in order to explore the nature of the receptor ligand(s) and the signals transduced to the pre–B cell. One of these antibodies, which is specific for the β-chain of the Ig receptor unit, may prove to be a universal B cell suppressant because it down-modulates the antibody receptors on all B cells.

Immunodeficiency Diseases

The pathogenesis of three primary immunodeficiency diseases is being investigated. X-linked agammaglobulinemia features an arrest in pre–B cell differentiation, the precise nature of which is under study in affected boys. Other studies address the hypothesis that selective IgA deficiency (IgA-D) and common variable immunodeficiency (CVID) may represent polar ends of a clinical spectrum reflecting a single underlying gene defect. The cellular defect in both deficiencies consists of arrested B cell differentiation. Both disorders frequently occur in members of the same family, and the same MHC class III haplotypes are associated with IgA-D and CVID. Current studies focus on the C4A complement gene.

Structure and Function of RNA Polymerase II

Jeffry L. Corden, Ph.D.—Associate Investigator

Dr. Corden is also Associate Professor of Molecular Biology and Genetics at the Johns Hopkins University School of Medicine. He received his B.S. and Ph.D. degrees in biochemistry and biophysics from Oregon State University. His postdoctoral work was done with Pierre Chambon in the Laboratoire de Genetique Moleculaire des Eucaryotes at the Faculte de Medecine, Strasbourg, France.

THE first step in the expression of genetic information is carried out by RNA polymerase II. How this enzyme determines which genes are to be expressed in which cells and under which physiological conditions remains largely unknown.

RNA polymerase II contains more than 10 different proteins whose precise functions are only now beginning to be addressed. The goal of my laboratory is to understand the structure and function of the subunits of RNA polymerase II. Several years ago we began to analyze the largest subunit of the mouse RNA polymerase II complex. This subunit comprises one-half of the mass of the enzyme and, through the work of many laboratories, is known to be involved in the enzymatic synthesis of RNA. We have cloned and sequenced the mouse gene encoding this largest subunit and have also isolated and characterized several mutations in this gene. These mutant genes are being used to study the function of different domains of the largest subunit. The aim of these experiments is to understand how RNA polymerase II orchestrates the ordered expression of 100,000 genes during the vertebrate life cycle.

The gene encoding the largest subunit of RNA polymerase II comprises 28 segments (exons) that cover 30,000 bases of mouse chromosomal DNA near the center of chromosome 11. The amino acid sequence deduced from the DNA sequence has revealed two interesting properties of the subunit. The major portion of the protein is similar in sequence to a bacterial RNA polymerase subunit (from *Escherichia coli*) that carries out an equivalent function. This evolutionary conservation is much stronger than had been expected and has allowed us to predict that the mouse subunit is involved in the transcription elongation process.

Although the major part of the largest subunit is related to the bacterial enzyme, our DNA sequence analysis has also revealed a domain that is unique to RNA polymerase II. This domain is located at one end of the molecule and constitutes a 52-fold repeat of a seven–amino acid sequence. This unusual sequence, although absent in bacteria, is found in other large subunits of virtually every RNA polymerase II, including those of animals, plants, insects, and protists. We are currently focusing our efforts on understanding the role of this domain in the process of transcription.

Our genetic approach to the function of this carboxyl-terminal domain (CTD) has grown from analysis of mutations in the largest subunit gene. We first isolated mutant mouse tissue-culture cell lines that are resistant to the mushroom toxin α-amanitin. The largest subunit genes from several of these cell lines have been cloned and, by reintroduction into amanitin-sensitive cells, have been shown to confer resistance to α-amanitin. We have used this gene transfer technique to map the mutations responsible for amanitin resistance.

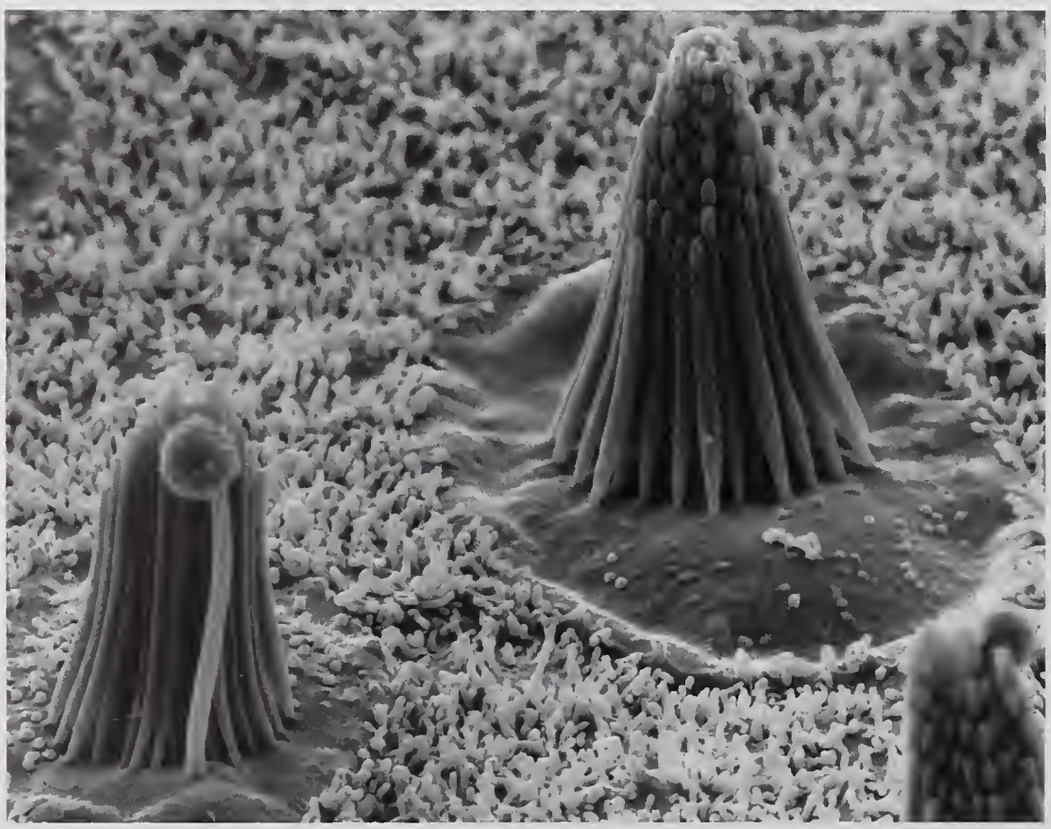
The availability of a well-defined, selectable genetic marker in the largest subunit gene has proved useful in the analysis of the role of the CTD. Deletion, insertion, and substitution mutations have been created in the CTD of an amanitin-resistance gene. The effect of these mutations has been tested by introducing the altered resistance genes into cells and scoring for amanitin resistance. Removing the entire CTD eliminates the ability to transfer amanitin resistance, demonstrating that the CTD plays an essential role in transcription. We have also shown that up to 20 of the 52 repeats are dispensable for growth in tissue culture, indicating either that the CTD is functionally redundant or that dispensable repeats are only necessary in some cell types. Paradoxically, we have recently shown that the CTD is highly conserved among all mammals. We are now designing experiments to assess the role of dispensable heptapeptide repeats during development.

We have also been examining postsynthetic modifications of the CTD. This domain is rich in amino acids (such as serine, threonine, and tyrosine) that can be modified by addition of phosphate groups. RNA polymerase II is a phosphorylated enzyme, but no function for phosphorylation is known. We have used synthetic

CTD peptide repeats as substrates to identify enzymes (CTD kinases) that carry out this phosphorylation. Two CTD kinases have been isolated, each containing protein kinase catalytic subunits previously identified in yeast as cell-division control proteins. Further cloning studies have revealed several more related mouse kinases whose role in CTD phosphorylation is currently under investigation. This investigation is supported by a grant from the National Institutes of Health.

Using the purified CTD kinases, we have mapped the sites of phosphorylation on the CTD. These sites, serines that precede proline residues, correspond to sites in other proteins that are modified in a cell cycle–dependent fashion. Current studies are directed toward examining the level of CTD phosphorylation at different times in the cell cycle. One consequence of CTD phosphorylation is that the CTD becomes greatly extended. The basis of this conformational change is also an area of investigation in our laboratory.

To define more precisely the amino acid sequences required for CTD function, we have devised a strategy to clone synthetic CTD repeats. This work is being done in the yeast *Saccharomyces cerevesiae*, where as few as 10 of the repeats are sufficient for viability. We have constructed mutant CTDs in which the residues identified as phosphorylation sites have been changed. In most cases these mutations are lethal, indicating that phosphorylation is essential for RNA polymerase II function. We are currently using yeast genetics to isolate suppressors of these mutations in an effort to identify proteins that interact with the CTD. Through characterization of such proteins we hope to determine the function of the CTD in RNA polymerase II function.



Hair cells from the bullfrog inner ear. A tuft of cilia, which sense mechanical stimuli, protrude from the apical surface of each cell. Deflections of the bundle toward the longer cilia are excitatory.

Research of David Corey and John Assad.

Mechanically Activated Ion Channels

David P. Corey, Ph.D.—Associate Investigator

Dr. Corey is also Associate Professor of Neuroscience at Harvard Medical School and Assistant Physiologist at Massachusetts General Hospital, Boston. He studied physics as an undergraduate at Amherst College, conducted research for a year at Harvard Medical School, and then entered the neurobiology program at the California Institute of Technology. His thesis work, with James Hudspeth, focused on mechanical transduction in auditory receptor cells. His postdoctoral work with Charles Stevens at Yale Medical School was on voltage-sensitive ion channels.

OUR laboratory is interested in how protein channels in cell membranes mediate the electrical activity of the brain. Such channels, which open and close to regulate the flow of minute amounts of electrical current into a cell, are intimately involved in the brain's information processing. They are important in detecting sensory signals such as light and sound, in the transmission of this information from the sense organ to the brain, and in communication from one brain cell to another. We are focusing primarily on ion channels in the sensory receptor cells of the inner ear, especially on the mechanism of activation of those channels that detect sound.

The sound-activated channels occur in a special type of inner-ear cell called a hair cell. These cells are named for a bundle of cilia that extends from their top surface and that tilts back and forth with each cycle of a sound wave. Moving the cilia opens and closes the channels. Over 10 years ago, it was found that the ion channels could open extremely quickly—in just a few millionths of a second—when the bundle was deflected. That speed rules out the kind of biochemical chain used by receptor cells of the eye and nose. We proposed instead that mechanical forces on the cilia could pull the channels open directly. Others have shown that this was true, and it was possible to measure the opening movement of channels when force was applied to them. But where are the channels, and what pulls on them?

Other workers, having explored around the bundle with a fine electrode, indicated that the channels were near the tips of the cilia. This was a surprising result, since a lever action of the cilia might focus forces near the bases. We have recently confirmed this, however, by putting a fluorescent dye inside the cell that reports the entry of calcium through the channels. When calcium is allowed to enter the cell, the fluorescent signal appears first at the tips, showing that the channels are there.

With attention focused on the tips of the cilia, a group in England discovered that the tips are linked by extremely tiny filaments, just 150 nanometers long, which they called tip links. Together these clues led to a beautifully simple theory for the operation of hair cells. Deflecting the bundle in one direction would stretch the tip links, and they would pull directly on the ion channels to open them and let ions into the cell. Moving in the other direction would relax the links and allow channels to close. However, no way to test the theory could be found.

This past year, we have been able to confirm it. We found that removing calcium from the fluid around the cilia, for just a few seconds, completely eliminated the tip links, as observed with either transmission electron microscopy or scanning electron microscopy. When we tested the mechanical sensitivity of the hair cells, by moving a single bundle and measuring the electrical response of the cells, we found that the same treatment eliminated the electrical response in a few tenths of a second. A further test indicated that the electrical response was lost specifically because the mechanical tension on the channels was gone. Thus tip links do convey the stimulus tension to the ion channels.

In addition to this transduction mechanism in these cells, there is an adaptation mechanism that enables them to be sensitive to extremely small displacements while retaining an ability to respond over a large range of stimuli. Hair cells have a very narrow range of sensitivity. A deflection of about a third of a micrometer—the diameter of one cilium—is sufficient to open all channels. We found some years ago, however, that a steady deflection that opens all channels is rapidly followed by the spontaneous closure of channels. To reconcile this with the context of the tip-links model, one must suppose that closure is a consequence of relaxing the tension on the channels. A further deflection to stretch the tip links does, in fact, reopen channels. Thus there seems to be a continuous adjustment system, acting to set the tension on the ion channels. Indeed, another group found that the stiffness of the bundle relaxed with the same time course as adaptation, in keeping with a tension adjustment scheme.

In contrast, experiments from a different group

suggested that adaptation is not a mechanical adjustment, but comes about through calcium entry through the transduction channels. Calcium inside the cell would close channels, to create a similar kind of negative feedback. It was clearly important to distinguish these models.

In the past two years, we have obtained good evidence for the mechanical adjustment hypothesis. A key feature of this hypothesis is that the putative adjustment of tension in the tip links would also be felt by the cilia themselves and would act to move the bundle. We found, first, that we could change the rate of adaptation by changing the voltage across the cell membrane. Because voltage affected the relaxing more than the tightening, we could then predict that positive voltage would increase the tension, and how much and how quickly. Judging from the activation of transduction channels, these predictions were borne out. Third, we could then predict how much the tension change should move the bundles. The predicted change was only about 100 nanometers—a quarter the wavelength of light—so we worked out a video microscopy system to measure the motion with a resolution of 5 nanometers. When the voltage was changed, the bundles did move, with the predicted amplitude and time course. Thus a quantitative theory that describes the adaptation also describes the movement, strongly suggesting that adaptation is mechanical.

How could tension on the tip links be adjusted? Others have speculated that the adaptation comes about by a movement of the points where the tip links are attached to the structure of the cilia. When the tip links are stretched to open channels, the attachment points might slip to allow the links to shorten. Conversely, if the tip links are relaxed, the attachment points move up to stretch the links to restore the resting tension.

This past year, we began to test this structural theory, by measuring the position of the attachment points before and after deflections that cause adaptation. Bundles were given calibrated deflections and then fixed with glutaraldehyde in the deflected position. Measurements from transmission electron micrographs suggest that the attachment points do move in response to deflection, although we still need to test whether the movement is quantitatively consistent with the adaptation.

What could move the attachment point? Some circumstantial evidence suggests that the "motor" molecule is like the myosin protein that causes the contraction of muscle: The stiff cores of the cilia are composed largely of actin, on which myosin moves in muscle; the motor molecule in these cilia moves in the same direction and at the same rate as myosin; and glass beads coated with muscle myosin do move on the actin cores of these cilia. We are now seeking further evidence as to whether the motor is a form of myosin.

Our ultimate aim is to describe each link in the mechanical chain from cilia to channels, in terms of the protein identity of the links, their biophysical properties, and their relationship to each other. Answers to these specific questions will contribute to the long-range goal of a comprehensive theory of mechanically activated channels, not only in the ear but in the many other cell types that display a mechanical sensitivity.

In the past two years, we have started to use some of what has been learned about ion channels to understand human disease. Hyperkalemic periodic paralysis is a dominantly inherited muscle disease that causes sporadic weakness or paralysis. Exercise or certain foods that raise the level of potassium in the blood can bring on a paralytic attack. Colleagues at Massachusetts General Hospital had found that the voltage-sensitive sodium channel in muscle was genetically linked to this disease, suggesting that a defect in the channel might be the cause. Last year we were able to show that the channels are in fact defective. In the presence of excess potassium, they fail to inactivate normally, which allows sodium to enter the muscle and (indirectly) causes the paralysis.

A puzzle, however, is that paralytic episodes are often preceded in patients by myotonia—an excessive contraction of the muscle. Could the same defect cause both myotonia and paralysis? This year we have used a toxin derived from sea anenomes, which mimics the genetic defect in sodium channels when applied to normal muscle. Toxin-treated muscle did, in fact, show classic myotonia, with both increased electrical activity and increased tension. Thus a single genetic defect can have a graded effect. Induced only slightly, it causes myotonia; induced to a greater extent by potassium, it causes paralysis. A computer model of electrical activity in muscle shows the same result with a single defect. Explaining the pathology at the molecular level gives hints for effective drug treatment of the disorder.

Genetic Regulatory Mechanisms in Cellular Differentiation

Gerald R. Crabtree, M.D.—Associate Investigator

Dr. Crabtree is also Associate Professor of Pathology at Stanford University School of Medicine. He received his B.S. degree from West Liberty State College, West Virginia, and his M.D. degree from Temple University School of Medicine, Philadelphia. He was a senior investigator at NIH before coming to Stanford University.

CELLS acquire their final differentiated function by a complex interplay between primary genetic regulatory events in the nucleus and interactions at the cell membrane. Building on concepts largely provided from studies on lower animals, our laboratory has been exploring regulatory mechanisms that help determine how mammalian cells differentiate to assume their normal functions.

Isolation of a Trans-acting Regulator of Homeodomain Protein Function

Several years ago, we identified a tissue-specific transcription factor, HNF-1α, that interacts with essential regions of the promoters of a large family of genes expressed in endodermally derived tissues. After purifying the protein and cloning its gene, we found that HNF-1 contains a homeodomain similar to that found in genes determining body form in insects. Curiously, the protein dimerizes through an amino acid sequence unlike that found in other homeodomain-containing proteins. This led us to look for a protein that might heterodimerize with it and hence diversify its regulatory capabilities. We found such a protein by screening a hepatocyte cDNA library at low stringency. This protein, HNF-1β, which is expressed in a partially overlapping group of tissues with HNF-1α, contains a dimerization and homeodomain similar to those of HNF-1α, but a different transcriptional activation domain.

Surprisingly, HNF-1α was only able to activate transcription in certain tissues, suggesting the existence of a second tissue-specific protein that regulated its function. Dirk Mendel began a search for such a protein by purifying HNF-1α by means that do not disturb hydrophobic interactions. HNF-1α copurified with an 11-kDa protein that participated in the formation of a tetrameric complex. The 11-kDa protein, which was cloned, expressed, and found to enhance the affinity of dimerization between HNF molecules, was named DCoH (dimerization cofactor for HNF-1). In both cases DCoH required the amino-terminal 32 amino acids that constitute the dimerization domain for binding to the complex consisting of two molecules of HNF-1α and two molecules of DCoH. Developmentally, the DCoH protein is expressed early (about day 9 or 10 after fertilization) in mice and is expressed in a group of tissues that do not always express HNF-1α or -1β, suggesting that other tissues such as the brain contain proteins that can bind to DCoH. When co-expressed with HNF-1α, DCoH enhances transcriptional activation by several hundredfold; we are presently investigating the mechanism by which this occurs.

T Lymphocyte Activation and Differentiation

T lymphocytes undergo two biologically and medically important types of differentiation. The first occurs in the thymus and generates cells capable of directing an immune response to nearly any antigen. However, the cells produced by the thymus that circulate in our blood are immunologically nearly inert. They acquire immunologic function as a result of a second process of cellular differentiation that takes about 10–14 days and produces T cells that coordinate the actions of other cells involved in the immune response by production of cytokines and cell-cell interactions. This differentiation pathway is initiated by a complex interaction between the T cell and an antigen-presenting cell. The essential requirement for a commitment to specialized function is a highly specific interaction between histocompatibility molecules, antigen, and the antigen receptor. This critical interaction is only effective when stabilized by transient nonspecific interactions based on intracellular adhesive molecules. Finally, lymphokines such as interleukin-1 and -6 that are the secreted products of the antigen-presenting cell are accessory signals necessary to initiate differentiation. These three requirements for the initiation of differentiation—a highly specific cell-cell interaction, a nonspecific adhesive interaction, and cytokines—are similar to the requirements for the cellular commitment to differentiate in other systems.

Because the interaction between the antigen-

presenting cell and a T lymphocyte is transient, all of the molecular events required for the decision to proceed down this 10- to 14-day pathway of cellular differentiation must occur in the short period during which the T cell and the antigen-presenting cell interact. Our laboratory is seeking an understanding of the molecular basis of this cellular decision.

Our approach to understanding the events that initiate this pathway was to begin by defining the molecules essential for the activation of genes that are known to be essential for T cell activation and differentiation. This approach led us to define two transcription factors that specifically receive signals from the antigen receptor and are responsible for activation of early genes such as interleukin-2. One of these proteins, nuclear factor of activated T cells (NFAT), is expressed almost exclusively in T lymphocytes, and its transcriptional activity is under rigorous control by the antigen receptor. Furthermore, its transcriptional activity is induced by the antigen receptor immediately before the activation of most early genes. We thus focused our attention on this transcription factor rather than others that are activated in many cell types and under many biologic circumstances.

Immunosuppressants Cyclosporin A and FK-506 Block Nuclear Translocation of the Cytosolic Component of NFAT

Cyclosporin A and FK-506 specifically inhibit T cell activation, a characteristic that underlies their clinical use as immunosuppressants to prevent transplant rejection. Although the mechanism of action of these drugs is unknown, they appear to work early during the commitment period for T cells. By acting early, they block the late functions of T cells and many of the functions of B lymphocytes and other hematopoietic cells that are directed by T cells. Other groups have found that these drugs bind and inhibit the function of *cis-trans* prolyl isomerases. These enzymes accelerate the folding of newly synthesized proteins. In studies with Stuart Schreiber (Harvard University), we have shown that the isomerase activity is not involved in the action of the cyclosporin and FK-506, but rather that an inhibitory complex formed between either cyclosporin A and cyclophilin or between FK-506 and FKBP blocks signal transduction. This inhibitory complex binds and blocks the activity of calcineurin, a heteromeric protein phosphatase also known as PP2B. This inhibition is likely to be related to the mechanism of action of the FK-506 and cyclosporin A, since immunosuppressive drugs form complexes that block the activity of calcineurin but inactive ones do not. We have found that the specific transcriptional activity of NFAT, but not its DNA-binding activity, is affected by cyclosporin A and FK-506. Mike Flanagan, Peter Kao, and Blaise Corthesy found that NFAT is a complex heterodimeric protein, one subunit of which is constitutive, T cell specific, and located in the cytoplasm of resting cells; the other subunit is located in the nucleus and is induced. The transcriptionally active protein forms when the cytoplasmic component translocates to the nucleus in response to stimulation through the antigen receptor of T cells. Cyclosporin A and FK-506 appear to function by inhibiting the translocation but do not interfere with the induction of the nuclear component. The inhibition of nuclear import is specific to the extent that the drugs do not block nuclear import of NF-κB. Recently, Peter Kao and Blaise Corthesy have purified the proteins and isolated the genes for the two subunits of NFAT. This should allow us to determine directly if calcineurin is involved in the signaling pathway initiating T cell activation.

The Mechanism of a Bacterial Transposition Reaction

Nancy L. Craig, Ph.D.—Associate Investigator

Dr. Craig is also Associate Professor of Molecular Biology and Genetics at the Johns Hopkins University School of Medicine. After receiving a bachelor's degree in biology and chemistry from Bryn Mawr College, she did graduate work on bacterial responses to DNA damage with Jeffrey Roberts at Cornell University, leading to her Ph.D. degree in biochemistry. She did postdoctoral research on the integration/excision cycle of the bacteriophage λ with Howard Nash at the National Institute of Mental Health. Before joining HHMI, she was Associate Professor of Microbiology and Immunology and of Biochemistry and Biophysics at the University of California, San Francisco.

DESPITE DNA's essential role in maintaining accurate genetic information, this molecule displays a surprising degree of plasticity. We now know that DNA rearrangements—i.e., the reorganization of DNA sequences through breakage, translocation, and rejoining reactions—mediate a wide variety of fundamental cellular processes. DNA rearrangements play an important role in the control of gene expression during development, the acquisition of new genetic elements such as viruses, the repair of damaged DNA, and the creation of genetic diversity. We are interested in understanding at the molecular level how DNA rearrangements occur and are controlled.

We are particularly interested in the type of recombination called transposition. In this reaction, a discrete DNA segment moves from one position in a genome into another, nonhomologous target position. This translocation may occur between different positions on the same chromosome, between chromosomes, or between chromosomes and extrachromosomal elements such as plasmids. Transposable elements have been identified in a wide variety of organisms. A notable consequence of transposition is the promotion of rapid and extensive genetic change. Transposon insertion results in the stable linkage of information encoded by the transposon with the target DNA. Transposon insertion into a gene will likely inactivate the gene, and insertion into DNA sequences that control the expression of nearby genes may inactivate or activate those genes. Probably because of its potential for profound influence, transposition is highly regulated.

Our research is focused on understanding the transposition of Tn7, a bacterial transposon with several unusual properties. Of particular note is Tn7's unusual target selectivity. Most transposable elements display little insertion-site selectivity, inserting into many different targets. By contrast, Tn7 inserts at high frequency into a single, specific site in the chromosomes of many bacteria. In *Escherichia coli*, the organism in which we study Tn7, this special target is called an attachment site and designated *attTn7*. When *attTn7* is unavailable, Tn7 resembles most other transposable elements, inserting at low frequency into many target sites.

Like most mobile DNA segments, Tn7 encodes the machinery that mediates its transposition from place to place. We have established that Tn7 encodes a surprisingly complex array of transposition proteins, and we have also identified the particular DNA sequences at the ends of Tn7 and at its insertion sites that are the actual substrates for transposition. The high-frequency insertion of Tn7 into *attTn7* is mediated by four Tn7-encoded genes, *tnsABC* + *tnsD*; Tn7 insertion into other random target sites, a low-frequency reaction, is mediated by a distinct set of Tn7-encoded genes, *tnsABC* + *tnsE*. Tn7 also encodes resistance to the antibiotics trimethoprim and streptomycin/spectinomycin. The ability of Tn7 and other mobile DNA segments that encode antibiotic resistance to insert into and thus be joined to plasmids that can, in turn, move among a variety of bacterial species underlies the rapid dissemination of antibiotic resistance among bacterial populations.

Our overall goals are to understand in molecular detail how Tn7 moves from place to place and how the frequency of this movement is modulated. We expect that understanding the macromolecular interactions that underlie Tn7 transposition will contribute not only to the understanding of DNA recombination but also to the understanding of other complex protein–nucleic acid transactions such as DNA replication, transcription, and RNA processing. We are using a variety of biochemical and genetic approaches to dissect Tn7 transposition.

The fundamental steps in transposition are the DNA cleavages that separate the transposon from its flanking donor DNA and the subsequent breakage of the target DNA and the joining of the transposon to the target DNA. Little is known in molecular terms about how such reactions occur. Our

HHMI-supported effort is focused on dissecting the mechanism of the DNA strand breakage and joining reactions during Tn7 transposition.

We previously developed a cell-free system for Tn7 transposition to *attTn7*. Tn7 inserts efficiently and specifically into *attTn7 in vitro* in the presence of four purified Tn7-encoded proteins: TnsA + TnsB + TnsC + TnsD. Thus these proteins participate directly in transposition. We have also established that Tn7 strand breakage and joining can be carried out by a subset of these proteins: TnsA + TnsB + TnsC. We now want to know which transposition protein carries out which particular step(s) in transposition and the chemical basis of these reactions. We are probing these questions by examining the ability of each transposition protein to individually execute a subset of the chemical steps that underlie the complete transposition reaction. Our strategies to reveal what may be usually suppressed activities include manipulating the reaction conditions and using DNA substrates derivatized in particular ways that we suspect may provoke strand cleavage and joining.

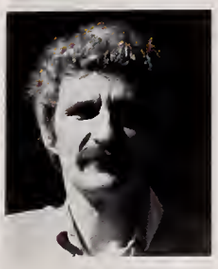
Another biochemical method we are using to identify the domains of the recombination proteins that are most intimately involved in DNA strand breakage and joining is to determine, through protein-DNA crosslinking studies, the segments of the recombination proteins that most closely appose the positions of DNA strand breakage and joining during transposition.

In a complementary genetic approach, we are seeking to identify "active sites" in the recombination proteins that promote DNA strand breakage and joining, by isolating and characterizing mutant transposition proteins altered in their ability to promote these reactions. We suspect that these active sites actually lie in TnsB, which binds specifically to the ends of Tn7, i.e., the sites of strand breakage and joining during transposition. We are also particularly interested in TnsB because there is some sequence similarity between TnsB and the recombinases of retrotransposons, including the integrases of retroviruses. Our long-term goals are to describe in detail the chemical steps that underlie DNA breakage and joining and to understand how the recombination proteins promote these reactions.

We are also interested in understanding how transposition is controlled and, in particular, in understanding the interplay between Tn7 and its bacterial host.

Much of our biochemical characterization of Tn7 transposition has focused on dissecting the high-frequency insertion of Tn7 into its specific insertion site *attTn7*. We are now working to extend the biochemical analysis of Tn7 transposition to its low-frequency insertion into random target sites. We are developing a cell-free system for Tn7 insertion into random target sites. We will then be poised to dissect the differences between high- and low-frequency Tn7 transposition in molecular detail. We also suspect that characterization of Tn7 insertion into random target sites will reveal host proteins that likely participate in this recombination reaction. We are also examining how the structure of the *E. coli* chromosome may influence Tn7 transposition. Another strategy we are using to probe the control of Tn7 transposition is to isolate and characterize mutant Tns proteins that display altered transposition properties. The above work on the control of Tn7 transposition is supported by a grant from the National Institutes of Health.

Mechanisms of Antigen Processing



Peter Cresswell, Ph.D.—Investigator

Dr. Cresswell is also Professor of Immunobiology at Yale University School of Medicine. He was born and educated in the United Kingdom. He received undergraduate degrees in chemistry and microbiology from the University of Newcastle upon Tyne and his Ph.D. degree in biochemistry and immunology from London University. Dr. Cresswell took postdoctoral training at Harvard University with Jack Strominger. Before assuming his position at Yale, he was Chief of the Division of Immunology at Duke University Medical Center.

FOREIGN protein antigens must be proteolyzed into peptides and must associate with specialized membrane glycoproteins before they can be recognized by T lymphocytes. These membrane glycoproteins are known as major histocompatibility complex (MHC) molecules, because of their original definition as the critical antigens responsible for organ graft rejection between members of the same species. Two types of MHC molecules have evolved, apparently to deal with two different categories of antigens. Class I MHC molecules bind peptides derived from cytosolic proteins synthesized by the cell that bears them, and are critical for T cell recognition of virally infected cells. Class II MHC molecules bind peptides derived from proteins internalized by a number of class II–positive cell types, such as B cells or macrophages, collectively known as antigen-presenting cells. The major current interest of my laboratory is in the molecular mechanisms involved in generating these MHC-peptide complexes.

Class I MHC-Peptide Association

The association of cytosolic protein-derived peptides with class I MHC molecules presents a topological problem, in that the peptides or their precursor proteins must cross an intracellular membrane for binding to occur. Evidence from the literature suggests that peptide binding occurs early in the intracellular transport of class I MHC molecules, perhaps in the endoplasmic reticulum (ER).

A mutant cell line, T2, derived in our laboratory from a similar cell line, .174 (produced by Robert DeMars at the University of Wisconsin–Madison), is defective in one or more of the steps involved in generating class I MHC-peptide complexes. Somatic cell genetic evidence suggested that the gene (or genes) involved is localized in a region of the MHC that is deleted in T2. Two of these genes are homologous to a group of proteins with multiple membrane-spanning domains known as the ATP-binding cassette (ABC) family of membrane transporters.

These molecules are generally involved in the active transport of small molecules, or occasionally larger proteins, across membranes. It has been proposed that the members of the family encoded in the MHC are responsible for the translocation of cytosolic peptides into the ER, where they bind to class I MHC molecules. Data from my laboratory have shown that most class I alleles expressed in T2 indeed lack peptides when they are affinity purified and associated peptides are analyzed chromatographically.

Human class I MHC (HLA) molecules generally fail to be transported to the cell surface when expressed in T2, a probable consequence of the lack of associated peptides. An exceptional allele, HLA-A2, is significantly surface-expressed at 20–50 percent of wild-type levels. HLA-A2 is associated with a limited set of peptides in T2, three of which have been isolated and sequenced. Two proved to be derived from the signal sequence of an interferon-inducible protein known as IP-30, and the third corresponded to no known protein.

The two identified peptides, a nonamer and 11-mer respectively, are presumably transported into the ER by the conventional signal-recognition mechanism involved in the translocation of secretory and transmembrane proteins. This would explain their association with HLA-A2 in a cell line defective in the normal mechanisms of peptide transport. In wild-type cells, class I–associated peptides are generally nonameric. The existence of the 11-mer peptide, and the fact that the third peptide is a 13-mer, argues that the generation of nonamers in wild-type cells may be a normal facet of peptide generation and transport.

Curiously, mouse class I MHC (H-2) molecules in general are well surface-expressed in the T2 cell line, even though they are devoid of associated peptides. This has led us to hypothesize that the human class I molecules are subject to a specific retention mechanism that prevents them from leaving the ER unless they are associated with a peptide. Understanding the differential transport properties of class I H-2 and HLA molecules in both normal and mutant cells is a major area of emphasis in our laboratory and should un-

cover additional details of the mechanisms involved in the *in vivo* generation of class I MHC-peptide complexes.

Class II MHC-Peptide Association

Class II MHC molecules associate late in transport with peptides derived from internalized protein antigens. It is believed that in the early stages of transport, they follow the same pathway as class I molecules, namely from the ER through the Golgi complex. This leads to the question, Why do class II molecules *not* bind peptides from the same set of cytosolically derived peptides that associate with class I MHC molecules?

At least a partial answer lies in the association of class II molecules with an additional glycoprotein, the invariant chain. This molecule is not a product of the MHC and is not structurally related to class I or class II glycoproteins. It associates with class II molecules immediately upon synthesis in the ER, forming a large nine-chain structure that we have found to consist of a core trimer of invariant chains associated with three class II molecules, each of which is a two-subunit heterodimer. Class II molecules in this structure cannot bind antigenic peptides, though they acquire the capacity to do so if released by mild denaturation or partial proteolysis of the invariant chain. Presumably their association with the invariant chain *in vivo* prevents class II MHC molecules from binding inappropriate peptides early in transport.

After assembly, the nine-chain class II MHC–invariant chain complex is transported from the ER and through the Golgi apparatus. We have found that the invariant chain then targets the complex to endosomal structures, where it encounters internalized foreign antigens, shown strikingly when influenza virus is used as a test antigen. Evidence obtained with proteinase inhibitors, such as leupeptin, suggests that the invariant chain is proteolytically degraded in the endosome, releasing class II molecules capable of binding peptides.

The simplest model for the generation of class II MHC-peptide complexes suggests that internalized proteins are endosomally degraded into peptides that bind to the newly released class II molecules. Recent evidence, however, suggests that this model is too simple. We have found that the T2 cell line, referred to above as a class I MHC antigen-processing mutant, is also a class II MHC antigen-processing mutant. T2 itself lacks class II molecules as a result of the homozygous deletion of the MHC region encoding class II molecules. However, if mouse class II molecules ($I\text{-}A^k$) are expressed in T2 by transfection, the cell line is unable to process internalized protein antigens into class II MHC-associated peptides recognizable by T cells. To date, it appears that both mouse and human class II molecules expressed in T2 associate normally with the invariant chain and are transported to an endosomal compartment where the invariant chain is proteolytically cleaved. The missing steps in generating functional class II MHC-peptide complexes remain to be determined, but again appear to be the properties of MHC-linked genes.

Regulation of Human Retroviral Gene Expression

Bryan R. Cullen, Ph.D.—Associate Investigator

Dr. Cullen is also Associate Professor in the Section of Genetics and the Department of Microbiology and Immunology and Associate Medical Research Professor in the Department of Medicine at Duke University Medical Center. He received his master's degree in virology from the University of Birmingham, England. After emigrating to the United States, Dr. Cullen worked as a research technician for several years before reentering graduate school at the University of Medicine and Dentistry of New Jersey, where he received his Ph.D. degree in microbiology. Before accepting his current position at Duke, Dr. Cullen studied gene regulation in higher eukaryotes, as a laboratory head in the Department of Molecular Genetics at Hoffmann-La Roche.

RETROVIRUSES derive their name from their ability to reverse the normal flow of genetic information from DNA to RNA. They possess the remarkable ability to synthesize a double-stranded DNA copy of their single-stranded RNA genome and then to integrate this DNA copy into the genome of the infected host cell. Once the genome of a retrovirus is integrated into a host chromosome, it is indistinguishable from a host gene and may be actively transcribed by the host's transcriptional machinery.

The infection of an animal by retroviruses can result in a number of disease states, of which the most common is leukemia. Indeed, the avian leukemia virus (ALV) was the first oncogenic virus to be defined experimentally. ALV and the somewhat similar murine leukemia viruses continue to be studied extensively as models for this virus group. This research has not only helped to delineate the retroviral replication cycle but has also greatly advanced our understanding of retroviral oncogenesis. Most importantly, this research has allowed the definition of a number of cellular genes, the oncogenes, whose inappropriate expression can contribute to cellular transformation.

Although animal retroviruses have been the subject of scientific research for some time, the discovery of human retroviruses occurred only within the last decade. Two major groups of pathogenic human retroviruses have been identified thus far. Human T cell leukemia viruses (HTLV-I and HTLV-II) are known to be causative agents of human leukemias, including adult T cell leukemia, and are significant disease agents in several parts of the world, including Japan and the Caribbean basin. Of even more concern are the human immunodeficiency viruses (HIV-1 and HIV-2), which are a leading cause of disease and death in parts of Africa and in the United States. A third group of retroviruses, the human foamy viruses (HFV), has recently been detected in human populations in both Africa and Oceania but has not yet been clearly associated with any disease.

A striking feature of all three classes of human retroviruses is that they each encode regulatory proteins that control both the quantity and quality of viral gene expression. This regulatory complexity is not observed in many animal retroviruses, including the avian and murine leukemia viruses, and may strongly influence the pathogenic potential of these "complex retroviruses." The major focus of this laboratory has been the determination of the role and mechanism of action of these viral trans-activators, with a concentration particularly on the Tat and Rev regulatory proteins of HIV-1.

In the past we demonstrated that the Tat protein of HIV-1 acts on sequences located within the HIV-1 promoter element (the long terminal repeat or LTR) to increase the level of expression of linked genes. This increased viral gene expression occurs via a bimodal mechanism that involves an increase in the rate of transcription of HIV-1 mRNAs and in the efficiency of translational utilization of those RNAs. The target sequence for Tat is a 59-nucleotide RNA stem-loop structure located at the very 5′ end of all viral mRNA molecules. The direct interaction of Tat with this RNA structure leads to an enhancement of viral transcription initiation and elongation. This mechanism, which may be unique to Tat, remains poorly understood but is likely to involve the interaction of Tat with cellular proteins that are currently poorly defined. The identification and characterization of these cellular proteins is a major research aim of this laboratory.

A second HIV-1 protein, Rev, is required for the expression of viral structural proteins but is dispensable for the expression of viral regulatory proteins. Our research has demonstrated that Rev acts post-transcriptionally to induce the cytoplasmic expression of the unspliced or incompletely spliced RNAs that encode the viral structural proteins Gag and Env, while simultaneously repressing the expression of the fully spliced RNAs that encode the viral regulatory proteins, including

Rev itself. The Rev protein therefore regulates its own expression via a negative feedback mechanism. It has been proposed that Rev achieves this effect by specifically regulating the export of viral RNAs from the cell nucleus to the cytoplasm. This specificity is conferred by a cis-acting viral RNA target sequence, the Rev response element (RRE), which has been shown to form a complex RNA secondary structure. Recent data demonstrate that Rev specifically recognizes, and binds to, a short, approximately 13-nucleotide primary sequence within the context of the larger RRE structure. Rev function also appears to require the subsequent binding of additional Rev protein monomers to secondary target sites within the RRE.

Mutational analysis of the Rev protein has demonstrated the existence of two functional domains. The first is a sequence-specific RNA-binding domain required for binding to, and multimerization on, the RRE, while the second is believed to interact with a currently unidentified cellular protein that may form part of the cellular RNA transport machinery. Mutations of this latter domain, the Rev activation domain, give rise to Rev proteins that act as competitive inhibitors of the wild-type Rev trans-activator. Mutant HIV-1 proteins of this type (dominant negative mutants) may have future application in the gene therapy of HIV-1–infected individuals. A major focus of this laboratory is the development of these trans-dominant Rev mutants and, in particular, the further investigation of the role of cellular proteins in the Rev response.

Finally, we have begun to expand our research to other human retroviruses, including HIV-2 and the apparently nonpathogenic HFV, as well as to related animal retroviruses, such as visna virus. The elucidation of similarities and differences in the regulation of gene expression among these retroviruses should facilitate the identification and understanding of the cis- and trans-acting elements required for their replication and pathology.

Mechanism of Retrovirus Infection

James M. Cunningham, M.D.—Assistant Investigator

Dr. Cunningham is also Assistant Professor of Medicine at Brigham and Women's Hospital and Harvard Medical School. He received a B.S. degree in chemistry from the University of Michigan and an M.D. degree from Stanford University School of Medicine. After clinical training in internal medicine (Peter Bent Brigham Hospital) and oncology (Dana Farber Cancer Institute), he was a postdoctoral fellow in the laboratory of Robert Weinberg at the Massachusetts Institute of Technology. Dr. Cunningham was an HHMI Associate at Brigham and Women's Hospital and Harvard Medical School before assuming his current appointment.

VIRUSES are parasites. They cannot produce progeny on their own, but must rely on the machinery provided by the host cell to replicate the viral genome and assemble new virus particles. Infection is initiated by attachment of the virus to the host cell—the first step in a complex reaction that results in transfer of the viral genome through the cell membrane and into the cytoplasm. This attachment, or binding, is a consequence of the interaction between proteins exposed on the surface of the virus and the host cell plasma membrane.

Cells that do not express a suitable virus-binding protein, called a receptor, are not susceptible to infection by a particular virus. Indeed, the ability of many pathogenic viruses, such as human immunodeficiency virus, poliovirus, and certain herpesviruses, to infect specific host tissues has been closely correlated with the expression of specific receptors.

Our laboratory has been interested in the mechanism of infection utilized by Moloney murine leukemia virus (MoMuLV), a member of a group of related leukemogenic retroviruses found in virtually all vertebrates. We have isolated a molecular clone, MCAT, which confers MoMuLV infectivity upon introduction into mammalian cells that are not normally susceptible to infection. Subsequent experiments have demonstrated that MCAT encodes for a membrane protein that serves as the MoMuLV receptor. Our current research is addressed toward dissecting the molecular details of the virus-receptor interaction that mediates infection and understanding the function of the receptor in normal cell metabolism.

The MoMuLV receptor is not present in mice for the convenience of the virus, but rather provides a portal for entry of lysine, arginine, and ornithine, amino acids that carry a net positive charge. There is a similarity between MCAT and two membrane proteins in yeast that are also amino acid transporters. This suggests conservation of a single mechanism for transport of these amino acids over evolutionary time and predicts that proteins similar to the MoMuLV receptor are used by all animals. Inherited disorders of cationic amino acid transport have been described in patients that may be explained by mutations in MCAT genes, a hypothesis we are now examining.

A protein that is closely related to the MoMuLV receptor has been identified in liver tissue. We have demonstrated that this protein is also an amino acid transporter with properties that may help explain the specific requirements for arginine metabolism by the liver: the capacity of hepatocytes to clear the portal vein of the high concentration of amino acids present after a big meal and the regulation of the urea cycle, the metabolic pathway used to eliminate nitrogen waste. Related forms of this protein are expressed in T and B lymphocytes and macrophages that are activated as part of the host response to bacterial and parasitic infection. Within the past few years, arginine has been identified as the substrate for nitric oxide, an important mediator of the host defense against these pathogens. Currently we are investigating how the family of related amino acid transporters can influence nitric oxide production by regulating arginine availability. As part of these studies, we have identified a molecular clone that encodes nitric oxide synthase, the enzyme that is responsible for the synthesis of nitric oxide from arginine. Expression of the gene encoding this enzyme is stimulated in macrophages that are activated in response to infection.

Our laboratory remains interested in how retroviruses interact with MCAT to gain entry into the cell. Recently we have prepared an antibody that can recognize the MCAT protein and thereby permits examination of its synthesis in infected and uninfected cells. These studies have demonstrated that MCAT is normally modified by the addition of carbohydrate during its transit from the ribosome to the plasma membrane. In cells that have been infected with MoMuLV, we have identified an interaction between MCAT and the virus envelope protein that impairs MCAT maturation and decreases amino acid uptake. This decrease may result from failure of the mature MCAT protein to arrive at the plasma membrane. This finding has important implications for un-

derstanding the host response to retrovirus infection and also for understanding virus interference —the production by cells of retroviruses that are resistant to additional infection by the same virus. To investigate this problem, we are now examining the behavior of MCAT proteins that contain mutations at the site of virus attachment. Our long-term goal is to understand the chemical basis of the virus-transporter interaction in sufficient molecular detail to design small molecules that can block virus binding and prevent infection.

The Nuclear Pore Complex

Laura I. Davis, Ph.D.—Assistant Investigator

Dr. Davis is also Assistant Professor in the Section of Genetics and the Department of Cell Biology at Duke University Medical Center. She received her undergraduate degree from the University of California, San Diego, and her Ph.D. degree from the Rockefeller University, where she studied with Günter Blobel. Her postdoctoral work was done with Gerald Fink at the Whitehead Institute.

ONE of the hallmarks of eukaryotic cells is the presence of membranous barriers that divide the cell into functional compartments called organelles. How these compartments are generated and how traffic through them is controlled are matters of intense interest in cell biology. Our focus is on the nuclear envelope, which separates the genomic material from the rest of the cell. Both RNAs and proteins move across the envelope, between the nucleus and cytoplasm, and the populations found in the two compartments are very different.

For example, newly synthesized mRNA must be extensively edited or processed before leaving the nucleus. The mechanism that restricts its export until processing is complete remains unclear. Similarly, only certain proteins are imported into the nucleus. These contain specific signals that must be recognized by some component of the import apparatus. Recently it has become clear that the cell can regulate the availability of these signals by changing the context in which they are found. For example, transcriptional activators that control the growth state of the cell by turning on gene expression can be held in an inactive form in the cytosol if their signals are hidden from the import apparatus. When the cell is stimulated to begin growth, the signals are uncovered, allowing the transcriptional activators to enter the nucleus and exert their function. Thus nucleocytoplasmic transport has an important regulatory role in controlling gene expression.

It is thought that all nucleocytoplasmic transport proceeds through large proteinaceous channels that perforate the nuclear envelope. Called nuclear pore complexes (NPCs), these structures are about 30 times as large as an ion channel and are probably composed of over 200 different polypeptides. Among the few known components of the NPC are members of a family of related proteins called nucleoporins. These proteins have been localized to an iris-like structure in the middle of the NPC, called the central transporter, and are thought to play an essential role in mediating protein (and perhaps RNA) transport. They are required for binding of proteins to the NPC prior to nuclear import, and antibodies that bind to the nucleoporins can block the energy-dependent movement of proteins through the NPC. While it is possible that the nucleoporins themselves recognize the signals on proteins destined for the nucleus, most of the available evidence suggests that they are more likely to act as docking points for cytosolic receptors that actually recognize the signals.

Our goal has been to understand more about the function of the nucleoporins and to identify other components of the transport apparatus. We use the budding yeast *Saccharomyces cerevisiae*, since it is amenable to both genetic and biochemical analysis. Using antibodies that recognize the mammalian nucleoporins, I previously identified homologues of these proteins in yeast and cloned the gene encoding one of them (*NUP1*). The genes encoding two others have now been isolated (*NUP2* and *NSP1*). The protein encoded by *NUP1* is essential for viability, and we have isolated *nup1* mutants that confer growth at one temperature but not another. These conditional mutants have been used to assay the phenotype resulting from loss of *NUP1* function. Using immunofluorescence, we have found that protein import into the nucleus is severely inhibited at the nonpermissive temperature, providing *in vivo* evidence that NUP1p is required for protein transport.

Conditional mutants will also be useful for identifying proteins that functionally interact with NUP1p. To do this, we are using genetic screens to find mutations in new genes that either enhance or suppress the conditional phenotype of the *nup1* mutants. The assumption is that mutations in proteins that functionally interact with NUP1p will either exacerbate the defect in *nup1* mutants and lead to death at all temperatures, or will overcome the defect and allow growth at the restrictive temperature. Using this approach, we hope to identify new NPC components, as well as cytoplasmic and nuclear proteins that may interact transiently with the nucleoporins in a functional manner.

We have also taken a biochemical approach to

identifying proteins that interact with NUP1p. Using a specific antibody, we isolated the *NUP1* protein from cell extracts and have found it to be stably associated with several other proteins. Some of these appear to be nucleoporins, while others are as yet unidentified. We are now purifying the individual members of the complex, in order to obtain protein sequences that we can use to isolate the genes encoding them. We expect some of these proteins to correspond with those we identify through genetic interactions with *NUP1*.

Molecular Approaches to T Lymphocyte Recognition and Differentiation

Mark M. Davis, Ph.D.—Investigator

Dr. Davis is also Professor of Microbiology and Immunology at Stanford University School of Medicine. He received his B.A. degree from the Johns Hopkins University and his Ph.D. degree in molecular biology from the California Institute of Technology. He held positions at NIH as a postdoctoral fellow and Staff Fellow before joining the staff at Stanford. He is the recipient of the Eli Lilly award in Microbiology and Immunology and a Gairdner Foundation award.

WE have focused on several major areas in immunology that generally resolve into two questions: How do T cells recognize foreign entities? How is lymphocyte differentiation controlled, both in the thymus and in the periphery? An additional goal is to refine and better integrate recombinant DNA technology with other powerful techniques in immunology, as an approach to defining the function of unknown genes or poorly understood genes and their products.

Topology of T Cell Recognition

The work of many investigators over the years has shown that T cells, through their antigen receptor molecules, recognize fragments of foreign proteins (peptides) embedded in major histocompatibility complex (MHC) molecules. In contrast, antibodies, although closely related to T cell antigen receptors (TCRs), bind intact foreign particles directly. This suggests fundamental differences in the rules governing T cell recognition versus antibody-mediated recognition by B cells. Because of the consistently high concentration of sequence diversity in the V-J junctional region of TCRs (equivalent to the third complementation-determining region [CDR3] of immunoglobulins) as well as structural considerations, we have proposed that this is the important region for peptide recognition and that other V region–encoded residues might contact the surface of the MHC molecule.

To test this hypothesis, we recently developed an immunological version of the classical genetic technique of second-site suppression. In these experiments we change residues that are important for T cell recognition (and not MHC interaction) on a peptide, immunize mice, and then analyze the responding T cells that emerge with respect to their TCR sequences. To hold part of the original receptor constant, we use mice that are transgenic for either chain of the original TCR. With this approach, we have shown that one of the two residues on the peptide that are important in the T cell response is governed by the CDR3 of the Vα polypeptide, and that the other (three amino acids downstream on the peptide sequence) is specified by the CDR3 of the Vβ polypeptide. Thus we have generated significant support for the original hypothesis. The results indicate that T cell recognition is a much more stylized event than antibody-antigen interactions.

Kinetics of T Cell Recognition and Activation

To learn more about the dynamics of TCR-peptide-MHC interactions, we have developed expression systems that allow us to produce either TCR or MHC class II heterodimers in a soluble form. This involves replacing the normal membrane-spanning sequences of these polypeptides with a signal sequence for lipid linkage, such as employed normally by a number of cell surface proteins. Molecules expressed in this fashion can then be conveniently cleaved from the surface of expressing cells with the enzyme phosphatidylinositol-specific phospholipase C. By utilizing high-density mammalian cell culture machines, we are able to make milligram quantities of a soluble TCR and its cognate MHC molecule. This has provided us with the raw material to initiate structural studies such as x-ray crystallography and nuclear magnetic resonance analysis. We have also used the soluble MHC protein to show greatly enhanced uptake of antigenic peptides at low pH. This is important both in making significant quantities of a pure antigen-MHC complex and in understanding the biology of this type of MHC molecule (class II) that recycles through low-pH endosomal compartments. Our current data suggest that the low pH triggers a specific conformational change in the MHC molecule, which allows it to bind new peptides more easily.

Recently we have studied the kinetics of TCR peptide-MHC interactions, using soluble MHC-peptide complexes that competitively inhibit binding of labeled anti-TCR antibody fragments to TCRs. In several cases studied, we derive K_d values for this interaction of 5×10^{-5} M, 1,000- to 10,000-fold weaker than antibodies to protein ligands of comparable size. These values are consistent with the scanning nature of T cell recogni-

tion and indicate that TCR binding to peptide-MHC ligands is so weak energetically that other, antigen-independent receptor-ligand systems must govern the initial stages of T cell contact with antigen-presenting or target cells. The best candidates for such regulators of T cell interaction are adhesion molecules, such as LFA-1 (lymphocyte function antigen 1) or CD2. This would give adhesion molecule–ligand interactions a major role in orchestrating which T cells interact with which antigen-presenting cells. This has important implications for autoimmunity: normally T cells would be expected to focus on appropriate antigen-presenting cells (such as B cells or macrophages) and would be hindered in surveying most other cells or tissues (which would lack the appropriate ligand expression).

Generation of Memory T Cells *in Vitro*

A hallmark of the vertebrate immune system is its ability to respond much more strongly and quickly to a second encounter with an antigen. Although some of this effect is due to the clonal expansion of B and T cells, a significant, and perhaps the major, contribution is thought to involve the induction of antigen-specific memory lymphocytes. Evidence suggests that this differentiation step is crucial to mounting a successful immune response, both in appropriate and inappropriate (e.g., autoimmune) circumstances. Studies of B or T lymphocyte memory have been hampered by the lack of well-defined systems in which uniform populations of primary and memory cells can be studied. We have approached this problem in T lymphocytes by making use of TCR $\alpha\beta$ transgenic mice that have an essentially monoclonal T cell immune system, thereby eliminating the clonal expansion component of T cell memory and allowing us to focus on the physiological changes that characterize this transition.

We find that primary T cells can only produce interleukin-2 (IL-2) in response to TCR stimulation with immobilized peptide-MHC complexes if PMA (phorbol 12-myristate 12-acetate) is included in the culture. In contrast, secondary, or memory, T cells, derived from the identical starting population, readily produce lymphokine on this substrate, irrespective of PMA addition. Both T cell populations are inhibited by protein kinase C (PKC) inhibitors H7 and staurosporine. Thus secondary T cells are functionally distinct from primary cells; the establishment of T cell memory, which may include alterations in PKC signaling, followed by the addition of exogenous IL-2, is sufficient to convert primary T cells to secondary cells *in vitro*. This provides an excellent opportunity to characterize and manipulate this important juncture in T cell differentiation.

Signal Transduction by the Epidermal Growth Factor Receptor

Roger J. Davis, Ph.D.—Assistant Investigator
Dr. Davis is also Associate Professor in the Program in Molecular Medicine and the Department of Biochemistry and Molecular Biology at the University of Massachusetts Medical School. He received his undergraduate and graduate education at Cambridge University and was a postdoctoral fellow with Michael Czech at the University of Massachusetts.

CELLULAR proliferation is a highly regulated process. During embryonic development, rapid cell growth is required to form the tissues of the body. In contrast, cellular proliferation in adults is slow, primarily serving to replace senescent cells. Adults, however, retain a limited capacity for rapid growth—for example, during wound healing. Regulation of this proliferative capacity is critically important. Errors in growth control can result in a variety of diseases, including cancer.

The local production of protein growth factors is an important mechanism that can account for the control of cellular proliferation. Our research group is investigating the action of a family of peptides that includes epidermal growth factor (EGF) and transforming growth factor-α (TGF-α). These agents are synthesized as cell surface glycoproteins that are split to release small soluble peptides. Both the membrane-bound precursor and the diffusible peptides are biologically active and bind to specific receptor molecules located at the surface of responsive cells. Secretion of these peptide growth factors contributes to the rapid growth of some tumors.

The long-term goal of this laboratory is to understand the molecular basis for the control of cellular proliferation by the EGF receptor. It is known that the binding of growth factors to this receptor at the cell surface triggers a complex series of chemical reactions that lead to DNA synthesis within the nucleus and to cell division. However, the molecular details of the signaling pathways utilized by the receptor are poorly understood.

Regulation of EGF Receptor Function

The EGF receptor is a glycoprotein consisting of an extracellular domain that binds growth factor, a membrane-spanning domain, and a cytoplasmic domain. The cytoplasmic domain is an enzyme, tyrosine kinase, that causes the covalent attachment of phosphate to tyrosine components of substrate proteins (phosphorylation). The binding of EGF to the receptor's extracellular domain causes an increase in the tyrosine kinase activity of the cytoplasmic domain. EGF also causes the receptor to aggregate and to associate transiently with intracellular regulatory molecules to form a signal transduction complex. We are studying these interactions and investigating the consequences of the phosphorylation process.

The ability of EGF to increase the tyrosine kinase activity of its receptor is blocked when cells are incubated with a tumor promoter or with other growth factors. Under these conditions, the EGF receptor is itself phosphorylated at multiple serine and threonine residues. We are investigating the significance of this phosphorylation. Our approach is to construct receptors with point mutations at the sites of phosphorylation, using recombinant DNA technology. These studies have demonstrated that the phosphorylation of a single threonine residue blocks the ability of EGF to stimulate the receptor's tyrosine kinase activity. Phosphorylation also alters the internalization of the receptor. We are investigating the structural basis for the effects of phosphorylation on the regulation of EGF receptor function. These studies are supported by a grant from the National Institutes of Health.

Signaling by the EGF Receptor

A principal question that we must answer to understand the mechanism of signal transduction by the EGF receptor is how a signal that is initiated at the cell surface can be transmitted to the nucleus to cause DNA replication. One class of regulatory molecules that could account for this process is the protein kinases. We are focusing our research on one subclass of these enzymes that cause the phosphorylation of proline-rich target sequences in substrate proteins. Examples are represented by the mitogen-activated protein kinases and the cyclin-dependent protein kinases. Each of these types of protein kinases exists in multiple forms as part of an extended family of enzymes that are regulated by growth factors and by the cell cycle. We are investigating the structure of additional members of this family by molecular cloning.

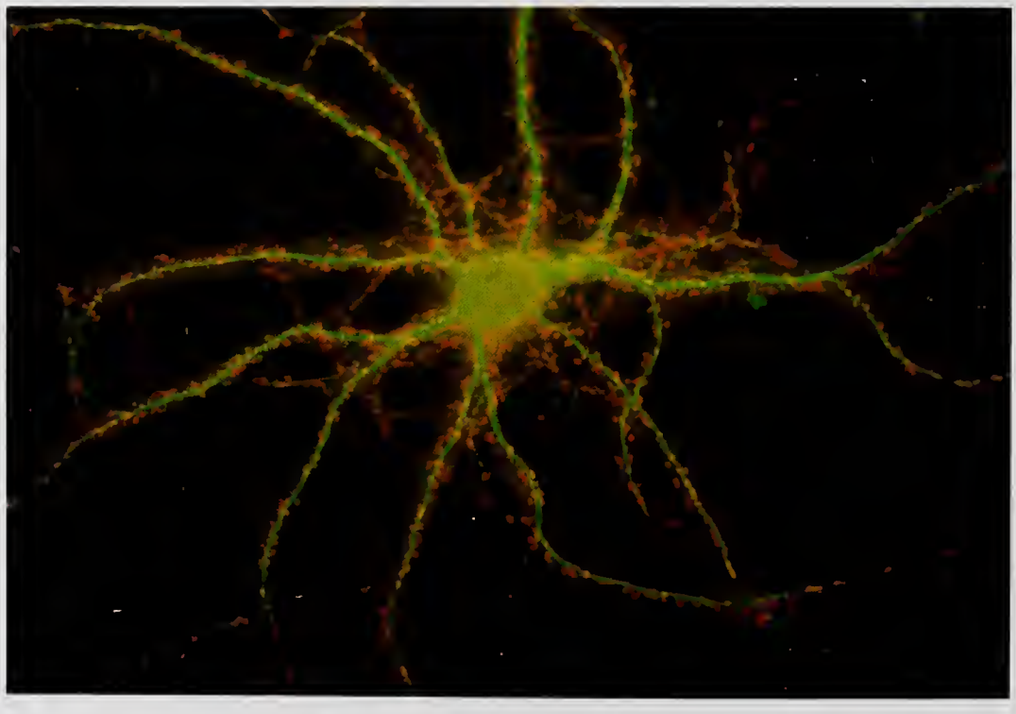
Substrates for these growth factor–regulated

protein kinases that we have identified include the EGF receptor and the nuclear transcription factors expressed by the proto-oncogenes c-*myc* and c-*jun*. We are investigating the mechanism and significance of the phosphorylation of these proteins in EGF-treated cells. The overall goal of these studies is to establish the molecular details of a signal transduction pathway that is initiated at the EGF receptor and leads to the regulation of nuclear function.

Tissue Specificity of Tumor Induction

The gene for the EGF receptor is a frequent site of integration by avian leukosis viruses. Insertion of the virus causes the expression of a truncated EGF receptor. The formation of a virus containing a copy of the truncated receptor gene can also occur. This truncated gene is the dominantly active oncogene *erbB*. The primary disease associated with *erbB* is erythroblastosis. However, mutations in the carboxyl terminus of *erbB* that occur during viral replication cause additional tumors—fibrosarcomas and angiosarcomas. We are studying the molecular basis for the change in tissue specificity of the *erbB* oncogene.

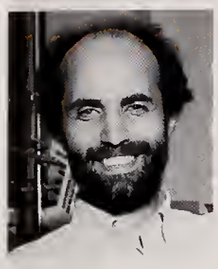
The approach we are taking is to construct recombinant viruses containing *erbB*. The advantage of this procedure is that we can undertake a systematic analysis of the effects of *erbB* mutations on the tissue specificity of tumor formation. The results should help explain why *erbB* causes tumors in one tissue but not in others. This information will be useful in designing strategies for clinical intervention in tumor development.



A hippocampal neuron in primary culture, showing axosomatic and axodendritic synapses at the surface. The perikaryon and the dendrites of the neuron are visualized by immunostain (green fluorescence) for MAP2, a protein associated with microtubules of perikarya and dendrites. Presynaptic nerve terminals originating from neurons not visible in the field are visualized by immunostain (orange-yellow fluorescence) for the synaptic vesicle protein synaptotagmin.

From the work of Michela Matteoli, Gary Banker, Thomas Südhof, and Pietro De Camilli.

Traffic of Synaptic Vesicle Proteins in Neurons and Endocrine Cells



Pietro De Camilli, M.D.—Investigator

Dr. De Camilli is also Professor of Cell Biology at Yale University School of Medicine. He received his M.D. degree from the University of Milano, Italy, where he worked with Jacopo Meldolesi. He did his postdoctoral studies with Paul Greengard at Yale University. Prior to his current position, he held appointments both at Yale and the University of Milano. Dr. De Camilli is a recipient of a McKnight Scholars Award and a member of the European Molecular Biology Organization (EMBO).

We are interested in the mechanisms of signaling between nerve cells, with emphasis on neuronal secretion of neurotransmitter molecules. Most neurotransmitters are stored in vesicles within nerve terminals and are secreted in response to depolarization of the terminal by fusion of the vesicle with the plasmalemma, a process called exocytosis. Most, and possibly all, nerve terminals contain two classes of secretory vesicles: synaptic vesicles (SVs) and the so-called large dense-core vesicles (LDCVs), which differ in a variety of properties, including neurotransmitter content.

SVs are small, morphologically homogeneous vesicles (50-nm diameter), containing nonpeptide neurotransmitters only. They are very abundant in nerve terminals, and clusters of SVs apposed to the plasmalemma represent a typical structural feature of axon endings. Their exocytosis takes place selectively at the presynaptic plasmalemma, is faithfully linked to nerve terminal depolarization, and plays the dominant role in the fast, point-to-point intercellular signaling typical of the nervous system (the so-called synaptic transmission). SV membranes are rapidly reinternalized after exocytosis and are used to re-form SVs, which are loaded locally with neurotransmitters.

LDCVs are larger vesicles with a dense protein-rich core. Their contents are predominantly peptide neurotransmitters, though they may also contain amines. Their exocytosis is preferentially triggered by trains of action potentials and is involved primarily in a modulatory type of signaling. They have long been recognized as the neuronal organelles equivalent to secretory granules of endocrine cells, which secrete peptide hormones and amines. In contrast, many of the properties of SVs set them apart from other secretory organelles. The elucidation of these properties represents the main focus of our laboratory. We are following three major lines of research. (Several aspects of this research have previously been supported by the National Institutes of Health.)

Biogenesis and Traffic of Synaptic Vesicles

First, we are investigating the biogenesis and traffic of SVs, using *in vivo* and *in vitro* systems. The large body of information recently accumulated on the molecular structure of SVs has made available numerous molecular probes that can be used in these studies (some of which are carried out in collaboration with Reinhard Jahn [HHMI, Yale University] and Thomas Südhof [HHMI, University of Texas Southwestern Medical Center at Dallas]). We have developed an immunocytochemical assay to detect SV exocytosis. Using this assay on hippocampal neurons in primary culture, we have shown that SVs undergo a high rate of constitutive exocytosis and recycling in developing neurons before synaptic contacts are formed.

These findings suggest that synapse formation coincides with a reorganization of the exocytotic machinery (rather than with its *de novo* assembly) and that signaling via SV exocytosis may play an important role in nervous system development prior to synapse formation. We are working at developing assays, some cell free, to define specific steps of the exo-endocytotic pathway of SVs. Such assays aid greatly in elucidating the underlying molecular mechanisms.

Synapse-like Microvesicles in Endocrine Cells

Second, we are exploring the relationship of SVs to organelles found in nonneuronal cells. Do SVs represent the adaptation of recycling vesicles to organelles found in other cells? How is the exo-endocytotic recycling of SVs related to the plasmalemma-endosome recycling that operates in all cells and occurs, for example, in the recycling of receptors?

We and others have found that a variety of SV proteins are expressed at significant levels by peptide-secreting endocrine cells. Investigating the subcellular localization of these proteins, we have found that at least some of them (synaptophysin, protein p29) are concentrated on a population of microvesicles (synapse-like microvesicles, SLMVs) distinct from the secretory granules that store and secrete peptide hormones.

A first biochemical characterization of these

microvesicles has demonstrated that they are similar in composition to bona fide neuronal SVs. Consistent with these findings, SLMVs, like SVs, undergo exocytosis, endocytosis, and recycling. We have therefore explored the possibility that SLMVs may share with SV the property to store and secrete nonpeptide neurotransmitters. The model endocrine cell we have used for these studies is the insulin-secreting cell of the endocrine pancreas, also called pancreatic β-cell.

Previous work had suggested that the nonpeptide neurotransmitter GABA, the major inhibitory neurotransmitter in the brain, may play a paracrine role in the physiology of the endocrine pancreas. GABA was reported to be concentrated in pancreatic β-cells and to act on surrounding pancreatic endocrine cells. Work we have carried out so far has suggested that SLMVs of β-cells may indeed store GABA. These findings support the hypothesis that the neuronal secretory pathway involving SVs represents an amplification and an adaptation of a secretory pathway that is expressed in a more rudimentary form in peptide-secreting endocrine cells.

Such results, in addition to having important implications for the field of endocrine physiology, raise the possibility of using endocrine cells to answer questions on the life cycle of SVs that would not be feasible to address in neurons. Using endocrine cell lines as a model system, we have obtained evidence indicating that the biogenesis of SLMVs (and presumably of SVs) is effected by protein sorting from endosomal membranes.

Molecular Mechanisms of Exocytosis

Third, we are interested in elucidating the molecular mechanisms of SV exocytosis. The approach we are currently taking is based on the assumption that at least some basic feature of the exocytotic process may have been conserved in evolution. We are exploring the potential of yeast genetics to identify some of the proteins involved. A variety of temperature-sensitive yeast mutants defective in vesicle exocytosis (*sec* mutants), and therefore in growth, have been isolated in recent years. We are currently searching for gene products that rescue *sec* mutations in a functional assay, by transfecting yeast *sec* mutants with mammalian brain cDNAs.

Stiff-Man Syndrome and Diabetes

The presence of SV-like organelles in endocrine cells suggests that pathological processes affecting SV proteins (mutations, autoimmunity) might affect both the nervous and endocrine systems. These considerations led us to identify a possible link between stiff-man syndrome and insulin-dependent diabetes mellitus (IDDM). Stiff-man syndrome is a rare and severe disease of the central nervous system characterized by a progressive rigidity of the body musculature. In a series of studies that involved a collaboration with the University of Milano, we have obtained evidence for an autoimmune pathogenesis of stiff-man syndrome and identified the GABA-synthesizing enzyme, GAD, as the major autoantigen.

GAD is concentrated at the cytoplasmic surface of SVs of GABAergic neurons and of SLMVs of pancreatic β-cells. IDDM, which results from an autoimmune destruction of β-cells, was found to occur at high frequency in stiff-man syndrome patients. This observation then led us to hypothesize an identity between GAD and the 64-kDa autoantigen of IDDM previously identified by Steinunn Baekkeskov, Åke Lernmark, and their coworkers (Hagedorn Research Laboratory, Denmark). A collaboration between our group and that of Dr. Baekkeskov (now at the University of California, San Francisco) confirmed that the 64-kDa antigen is GAD.

These findings may lead to a diagnostic assay for IDDM and for preclinical stages of the disease. In addition, it opens the possibility of investigating molecular mechanisms of autoimmunity in IDDM.

During the next year the work on SV exocytosis and on GAD autoimmunity in stiff-man syndrome will be partially supported by the National Institutes of Health.

Three-Dimensional Structures of Biological Macromolecules

Johann Deisenhofer, Ph.D.—Investigator

Dr. Deisenhofer is also Regental Professor and Professor of Biochemistry, and he holds the Virginia and Edward Linthicum Distinguished Chair in Biomolecular Science at the University of Texas Southwestern Medical Center at Dallas. He was born and educated in Germany. His Ph.D. research in protein crystallography was done at the Max Planck Institute for Biochemistry, Martinsried, and at the Technical University of Munich. As a postdoctoral fellow and as a staff scientist at the Max Planck Institute, he continued his structural analysis of biological macromolecules by x-ray crystallography. He has received several honors for his structure analysis of a photosynthetic reaction center, including the 1988 Nobel Prize in chemistry, which he shared with Hartmut Michel and Robert Huber.

MY colleagues and I study the three-dimensional structures of proteins to understand their folding, structural stability, and function. We are particularly interested in protein-pigment complexes catalyzing photochemical energy conversion, energy transfer, and electron transfer, and in membrane-spanning and membrane-associated proteins.

Cytochrome b/c_1 complex

The b/c_1 complexes (also called ubiquinol–cytochrome c oxidoreductases) are integral membrane proteins that play crucial roles in photosynthesis and cell respiration. Their function in these fundamental processes is to oxidize quinols to quinones and to transfer electrons and protons through the membrane. The electrons go to cytochrome c, and the protons build up an electrochemical gradient across the membrane, which, for example, drives the synthesis of ATP.

Photosynthetic purple bacteria have the simplest b/c_1 complexes, consisting of only three or four different protein subunits with three heme groups and an iron-sulfur cluster as cofactors. Photosynthetic reaction centers and b/c_1 complexes cooperate in the bacterial inner membrane; b/c_1 complexes occur at a concentration significantly lower than that of reaction centers and are therefore more difficult to isolate. David Knaff's research group (Texas Technical University, Lubbock) succeeded in purifying the b/c_1 complex from *Rhodospirillum rubrum*. In collaboration with this group we have prepared this complex in milligram quantities with high purity. This preparation serves as the starting material for crystallization experiments.

Mitochondria of higher organisms have in their inner membranes b/c_1 complexes consisting of at least 10 different polypeptide chains, 3 of which are similar to those in the b/c_1 complex of purple bacteria. We collaborate with Chang-An Yu and his colleagues (Oklahoma State University, Stillwater), who produced large crystals of the b/c_1 complex from beef heart mitochondria. Preliminary diffraction experiments at the Cornell High Energy Synchrotron Source (CHESS) were encouraging, and we hope to begin a structure analysis at medium resolution (approximately 5 Å) in the near future.

DNA Photolyase

Light energy plays an important role in reactions other than photosynthesis. The molecular machinery that enables cells to repair DNA damaged by ultraviolet (UV) light includes an enzyme that uses light energy to drive the repair of one frequent type of damage caused by UV irradiation: the crosslinking of two neighboring thymine bases in a strand of DNA. Most of these crosslinks are in the form of cyclobutane rings connecting four carbon atoms, two from each thymine base. The enzyme DNA photolyase can locate and bind to such defects, and upon input of light of suitable wavelength (300–500 nm) cleave the carbon-carbon bonds between the bases, thus repairing the damage.

DNA photolyase has been found in prokaryotes, eukaryotes, and archaebacteria. Aziz Sancar and his colleagues (University of North Carolina at Chapel Hill) sequenced, overexpressed, and purified the enzyme from *Escherichia coli*. It consists of a single polypeptide chain of 471 amino acids and two cofactors—a flavin-adenine dinucleotide (FAD) and 5,10-methenyltetrahydrofolate. The FAD cofactor fully reduced to $FADH_2$ is essential for the enzyme's function; the folate acts as a light-harvesting antenna.

Despite significant problems with the enzyme's tendency to denature, we were able to crystallize DNA photolyase from *E. coli* in two crystal forms; both forms diffract x-rays to at least 2.8-Å resolution. We collected x-ray diffraction intensity data from one of these crystal forms; to solve the phase problem, experiments are under way to bind heavy-atom compounds to the protein in the crystal. We are also trying to crystallize the enzyme in complex with a substrate, a five-nucleotide piece of single-stranded DNA contain-

ing a crosslinked pair of thymine bases. These experiments have to be done in the dark or under yellow or red light, to prevent the enzyme from repairing and releasing the substrate.

Because light is an essential ingredient of the enzymatic reaction of DNA photolyase, the enzyme-substrate complex will be a suitable system to study the time course of the reaction, using the Laue technique. This technique employs the broad spectrum of x-ray wavelengths in the synchrotron's powerful beam to record within a very short time a sufficient fraction of a crystal's diffraction pattern for structural interpretation. In such an experiment the crystal is irradiated with a light pulse suitable to trigger the enzymatic reaction, and diffraction patterns are recorded at different times after the pulse. These data can provide snapshots of the structural rearrangements during the reaction and thus contribute to an understanding of the enzyme's mechanism.

Cytochrome P-450 Enzymes

P-450 enzymes are members of a superfamily of b-type heme proteins that catalyze hydroxylation of organic substrates using electrons from reduced nicotinamide adenine dinucleotide phosphate (NADPH) and molecular oxygen. They consist of a polypeptide chain of about 420 amino acids and a heme group as part of their active site. P-450s occur in organisms as different as bacteria and humans. Some of them play specific roles in biosynthetic pathways, such as formation of glucocorticoids, androgens, and estrogens from cholesterol; others are involved in detoxification of foreign substances by making them more soluble and so facilitating their excretion. However, in some cases hydroxylation of foreign substances creates highly efficient carcinogens; thus the effects of P-450s are not always beneficial.

P-450s occur in two classes, distinguished by the way in which electrons from NADPH are transferred to the heme group. Class I P-450s receive the electrons via a flavoprotein and an iron-sulfur protein; class II enzymes are reduced directly by a protein containing the cofactors flavin mononucleotide and FAD. The three-dimensional structure of only one class I P-450, the camphor hydroxylase P-450_{cam}, is known at atomic resolution.

To learn more about the structural variation within the cytochrome P-450 family, about differences between class I and class II enzymes, and about the spectrum of reactions catalyzed by P-450s, we collaborate with Julian Peterson and his colleagues (University of Texas Southwestern Medical Center at Dallas) to analyze the three-dimensional structures of a class I and a class II enzyme. The class I P-450 was isolated from *Pseudomonas*; it hydroxylates the hydrocarbon α-terpineol and was therefore named P-450_{terp}. The class II P-450 is the cytochrome part of an enzyme from *Bacillus megaterium* that is unique in combining a P-450 domain and a P-450 reductase domain in a single polypeptide chain of 1,066 amino acids. The cytochrome domain, P-$450_{BM\text{-}3}$, consists of the 471 amino-terminal amino acids; it was cloned and expressed at high levels in *E. coli*. Its functions include hydroxylation and epoxidation of long-chain fatty acids.

Well-ordered crystals of both P-450s have been grown. At CHESS, crystals of P-450_{terp} diffracted to about 2.0-Å resolution, and crystals of P-$450_{BM\text{-}3}$ diffracted to better than 1.5-Å resolution. Molecular replacement calculations, using the P-450_{cam} model, produced a plausible solution for the crystal structure of P-450_{terp}. Multiple isomorphous replacement with three heavy-atom derivatives allowed the calculation of an electron density map at 3-Å resolution for P-$450_{BM\text{-}3}$; this map is currently being interpreted. Model building and crystallographic refinement will result in accurate models of both proteins in the near future. We also plan to study substrate binding and structures of site-specific mutants.

Other Projects

In addition to the projects described above, we also work on the determination of the three-dimensional structures of several other proteins, including the catalytic domain of human HMG-CoA reductase, a key enzyme in the synthesis of cholesterol and a likely target for drugs; human synapsin I, a protein binding to synaptic vesicles and mediating their release; the small GDP-binding protein smgp25A from bovine brain; the DNA-binding protein myogenin; the SecA protein from *E. coli*, one of the key parts of the protein export system; mammalian phosphofructokinase; and ribonuclease inhibitor from pig liver.

The work on cytochrome b/c_1 complexes and on DNA photolyase is also supported by a grant from the Welch Foundation.

Molecular Mechanisms of Lymphocyte Differentiation

Stephen V. Desiderio, M.D., Ph.D.—Associate Investigator

Dr. Desiderio is also Associate Professor of Molecular Biology and Genetics at the Johns Hopkins University School of Medicine. He received his undergraduate degree in Biology and in Russian from Haverford College and his M.D. and Ph.D. degrees from the Johns Hopkins University School of Medicine in biochemistry, and cellular and molecular biology. After a postdoctoral fellowship at the Massachusetts Institute of Technology with David Baltimore, he returned to Johns Hopkins.

ONE remarkable feature of the immune system is its ability to recognize and respond to an extraordinary variety of foreign molecules, or antigens. This exquisite specificity is achieved through protein receptors, which bind tightly to specific antigens. Such receptors are found on the surfaces of two types of immune cells: B and T cells.

One type of antigen receptor is the antibody molecule, or immunoglobulin. The site on the antibody molecule that binds to a specific antigen is genetically encoded by multiple short segments of DNA. At the onset of development of the immune system, these DNA segments are located at separate places in the genome, but during the maturation of antibody-producing cells (B cells), segments are joined to form intact immunoglobulin genes.

In addition to antibody, another class of antigen receptor is found on the surfaces of cells that mediate cellular immunity (T cells). The T cell receptor's antigen-binding site, like that of the antibody molecule, is encoded by multiple DNA segments that are brought together during T cell maturation.

After their genes are assembled, the antigen receptor molecules are expressed at the surfaces of B and T cells. Here, several interactions combine to trigger cell division and further maturation. These include the binding of antigens and specialized hormones called lymphokines with their respective receptors.

Lymphocyte Activation

A major goal of our laboratory is to find out how specific antigens and growth factors trigger the activation of B and T cells. Biochemical evidence has long suggested that enzymes called tyrosine kinases might be intimately involved. These enzymes regulate the activities of other proteins, and of each other, by adding regulatory chemical groups, phosphates, to specific sites. The kinases involved in B and T cell activation have, for the most part, proved elusive. We therefore set out to find new tyrosine kinase genes that are expressed in B or T cells, and we identified seven. Two are expressed only in immune cells, and we have concentrated on those.

One of the genes resembles a growth-regulating gene called c-*src*. Unlike c-*src*, however, which is turned on in many different kinds of cells, this new gene is only turned on in B cells and their developmental precursors. Accordingly, we call the gene *blk*, for B lymphoid kinase. The *blk* gene is activated early in B cell development and is expressed along with several proteins that are known to transmit signals across the B cell membrane, including the antibody heavy chain.

When B cells develop into antibody-secreting or plasma cells, they no longer express these signal transmission proteins or *blk*. This pattern suggests that the product of the *blk* gene interacts with a receptor that spans the B cell membrane and that senses the presence of antigen or a specific growth signal. In the spleen, the protein encoded by *blk* is found specifically in those locations where resting B cells encounter antigen, reinforcing the idea that this kinase functions in the triggering of the B cell's immune response.

How might extracellular signals activate the *blk* kinase? The kinase is inactivated by addition of a phosphate group to a specific site, and is activated when the phosphate is removed. If the target site on the *blk* kinase is mutated so that it cannot accept the inactivating chemical group, the enzyme is locked into the "on" state. Cells that contain this "on" version of the *blk* kinase grow in an unregulated way. Having an activated version of the *blk* kinase has made it easier to identify its targets. We have found one target to be an enzyme called phospholipase C-γ (PLC-γ), a key component of the B cell activation pathway. Evidence from other systems suggests that PLC-γ is activated by tyrosine kinases. We speculate that activation of the *blk* kinase in B cells activates PLC-γ, initiating subsequent signaling events.

The hormone interleukin-2 (IL-2) plays a central role in the immune response by inducing the multiplication of T cells. The binding of IL-2 to its receptor turns on a tyrosine kinase of unknown identity. We have recently identified six new tyrosine kinase genes that are expressed in cells de-

pendent on IL-2 for growth. One of these genes, which we call *itk* (for IL-2–inducible T cell kinase), is expressed specifically in T cells. The *itk* gene encodes a tyrosine kinase that resembles *src* kinases, but from its sequence we predict that it will differ from those kinases with respect to its location in the cell and the regulation of its activity. IL-2 is known to regulate expression of a number of genes, including a gene for its own receptor. In cells that have been withdrawn from IL-2, *itk* is expressed at low levels. Soon after addition of IL-2, *itk* is strongly turned on in parallel with the IL-2 receptor gene, suggesting that *itk* functions in the IL-2 response.

Antibody Gene Rearrangement

The total number of immunoglobulin or T cell receptor gene segments is large, but when any particular immunoglobulin or T cell receptor gene is assembled, only a handful of segments are selected and joined. As a result, many different combinations of segments are possible. It is this shuffling of small bits of DNA that generates much of the diversity of the immune response.

By using artificial DNA molecules that rearrange after they are introduced into immature B cells, we have been able to outline the general features of antibody gene assembly, but the mechanism at the level of interactions between molecules is not known. The products of at least three genes are known to play an intimate role in rearrangement. Two of these genes, *RAG-1* and *RAG-2*, have been shown by others to activate somehow antibody gene rearrangement. A third gene encodes a protein that all cells use to help repair DNA damage and that also functions at a late step in antibody gene assembly. In addition to the products of these genes, our laboratory has found a protein—NBP—that binds specifically to a DNA sequence element required for rearrangement, suggesting that it may be a part of the recombinational machinery. Members of our laboratory are studying the products of the *RAG* genes and NBP, with the aim of understanding their functions and interrelationships.

Since NBP was isolated from calf thymus, we cloned the *RAG-1* and *RAG-2* genes of the cow and identified regions of identity among the genes from cow, mouse, and human. Based on the predicted protein sequences of these conserved regions, antibodies were raised that specifically bind to the products of the *RAG-1* and *RAG-2* genes. Using these antibodies, *RAG-1* and *RAG-2* proteins were identified in extracts of calf thymus and isolated. We found these proteins are distinct from NBP. Interestingly, the sizes of the *RAG-1* or *RAG-2* proteins are slightly different from those predicted by their gene sequences, suggesting that they are modified after they are made. We have gone on to find that both RAG-1 and RAG-2 are modified by addition of phosphate groups. As was discussed earlier, this type of modification often serves to regulate protein function. We are intrigued by the possibility that such is the case for RAG-1 and RAG-2.

Molecular Mimicry by Antibody Molecules

As a result of antibody gene rearrangement, the structures of potential antigen-binding sites vary almost endlessly. It has been a long-standing notion in immunology that the structure of an antigen could be mimicked by a special type of antibody—an antibody raised against another antibody that was originally elicited by the antigen of interest. We can draw an analogy for this idea from the art of casting. A sculptor creates a positive image, which is used to produce a negative image, the mold, which in turn is used to reproduce the final work. An essential difference between this metaphor and reality is that in casting, one object is shaped by another, while in the antibody response, the binding of antigen selects a complementary shape from a universe of preexisting ones.

We have tested this idea in collaboration with research groups in Baltimore and Paris (led by Mario Amzel of the Johns Hopkins School of Medicine and Pierre Ronco, INSERM, Hopital Tenon). Antibodies (Ab1) were produced against a hormone, angiotensin II. Then a second antibody (Ab2) was raised against the anti–angiotensin II antibody. A third antibody (Ab3) was in turn raised against Ab2. The Ab3 antibodies bound to angiotensin II just as well as the original Ab1 antibodies, suggesting that Ab2 could mimic the hormone's structure.

This was proved by analyzing the structures of Ab1 and Ab3. Remarkably, we found that the sequences of their antigen-binding sites were nearly identical. We then determined the three-dimensional structure of the antigen-binding site of Ab3 in complex with angiotensin II. We discovered that the atoms in critical contact with angiotensin II in this complex are also found in the sequence of Ab1. Thus both Ab1 and Ab3 were elicited by similar structures, one represented by the original antigen, angiotensin II, and the other by a surface feature of the Ab2 antibody that mimics angiotensin II.

Expression of the gene BK27 *in a* Drosophila *embryo with an extended germ band. This gene, which encodes a homeodomain and a Paired domain, exhibits prominent expression in the ventral labial lobe, marking the progenitor cells for the salivary gland placode. The staining seen in the head is localized in the subantennal region.*

Research and photograph by Susie Jun in the laboratory of Claude Desplan.

Transcription Control During Early Drosophila *Development*

Claude Desplan, Ph.D.—Associate Investigator

Dr. Desplan is also Associate Professor and Head of the Laboratory of Molecular Genetics at the Rockefeller University. He graduated from the Ecole Normale Supérieure de Saint Cloud in France. He received his D.Sc. degree from the University of Paris, working with Baty Moukhtar and Monique Thomasset at the INSERM on calcium regulation. He was a Maître de Conférence at the Ecole Normale Supérieure de Fontenay until he moved to the University of California, San Francisco, to work as a postdoctoral fellow with Pat O'Farrell on the functions of homeodomain proteins during early development.

ALL information required for the development of a fertilized egg into a complete organism is contained in its own genetic material, contributed by both parents, and in products provided by the mother as the egg is formed. Genetic studies on the fruit fly *Drosophila* have identified most of the genes involved in the process of pattern formation. Many of these genes appear to encode transcription factors and to share some common protein motifs, such as homeodomains or zinc finger domains. The zygotic genome responds to maternal organizing factors through a network of transcriptional regulators to set up the body pattern of the embryo. Many of the *Drosophila* developmental genes have mammalian homologues expressed during embryogenesis. Thus it is likely that the mechanisms uncovered in flies are of general significance for the development of multicellular organisms.

The goal of our laboratory is to understand the molecular mechanisms involved in the regulatory interactions among developmental genes. In light of the complexity of the system, a productive approach is, first, to characterize the molecular interactions *in vitro* and, second, to design *in vivo* systems to test models consistent with properties uncovered through the first approach. We are investigating the events leading to patterning in two major systems. One involves the establishment of the early anterior pattern, while the other is a structure-function analysis of a gene that controls later steps of development. Independent support will be sought for the research described in the first part of this report.

Interactions Between Maternal and Zygotic Genes

The maternal homeodomain-encoding gene *bicoid* is one of the few examples of a morphogen. A concentration gradient of its protein product appears to control the establishment of the expression pattern of the first set of zygotic genes, in particular the anterior gap gene *hunchback* (*hb*). Using transgenic flies that carry artificial genes containing combinations of binding sites for the products of *bicoid* and *hb*, we have shown that the establishment of the simple *hb* pattern requires not only *bicoid* but also the synergistic participation of *hb* itself. This *hb* autoregulation may help explain the sharp, precise boundaries of its expression domain. It also has evolutionary implications for the development of other organisms that do not possess *bicoid*.

The role of *bicoid* is not only to control *hb* expression but also to direct the expression of a set of newly discovered genes required for head formation. Using the same approach as described above, we have demonstrated that *bicoid* action is modulated by the maternal genes from the terminal *torso* group. These genes, which encode a cascade of kinases, appear to act post-transcriptionally on the *bicoid* protein (Bcd) to repress its action in the terminal region. This leads to the expression of the head gap genes as an anterior stripe, responsible for development of a specific head region. While *bicoid* is setting up their posterior margin, the *torso* group, preventing activation by *bicoid*, sets up their anterior margin.

We recently identified a gene that has such an expression pattern and encodes a homeodomain with a DNA-binding specificity related to that of Bcd. We are investigating the control of the new gene's pattern of expression, and are analyzing phenotypes in the head region for genetic deletions of the locus.

The demonstration of a direct interaction between *bicoid* and the maternal *torso* group is the first clue for understanding the formation of head structure. It also presents us with a unique opportunity to use a combination of biochemical and genetic approaches to address the nature and regulation of the post-transcriptional modifications of a transcription factor in the context of a developing organism.

Regulation of the Segmentation Gene *hb*

Molecular dissection of the *hb* promoter by several groups has allowed definition of the sites of action of *bcd*. Our *in vivo* studies have uncovered a more complex regulation of *hb*. As described above, *hb* appears to be involved in the establishment of its own early expression, in synergy with *bcd*. We have identified *in vitro*, in the

region of the *hb* promoter, an Hb protein–binding site potentially required for full anterior activation. Mutations of this site are being tested *in vivo* for their ability to respond to *hb* function. In another region of the *hb* promoter, we have identified a cluster of Hb-binding sites that may mediate the other instance of *hb* autoregulation at a later stage in development.

Interestingly, these two examples of positive *hb* autoregulation contrast with the general repressive nature of *hb*. We have obtained *in vivo* evidence that, depending on the promoter context, *hb* can act as a positive or a negative regulator of gene expression. Finally, the *hb* promoter is also regulated by the neighboring gap gene *Krüppel*, for which two putative binding sites have been identified. This description and the experiments presented above may allow us to reconstruct a functional *hb* promoter made exclusively of minimal response elements identified for the regulators of *hb*.

DNA-binding Specificity of the Homeodomain

Most developmental processes involve genes that encode a homeodomain (HD). The HD includes a domain similar to the helix-turn-helix motif present in many prokaryotic DNA-binding transcriptional regulators. Our analysis of the function of the HD has led us to propose that the specificity changes among classes of HDs are due to the nature of a single amino acid at position 9 of the recognition helix. This position is not critical in the prokaryotic helix-turn-helix proteins, and indeed structural analysis of HD-DNA complexes recently showed that the HD-DNA interaction involves a new mode of recognition. In this mode, amino acid 9 fits in the major groove of DNA, in close contact with the base pairs being recognized.

We have generalized the role of amino acid 9 to several classes of HDs. Using a powerful selection procedure from a library of random sequences, we have isolated specific sequences recognized by a set of mutant HD proteins carrying different residues at position 9. These sequences confirm our previous observations, but indicate that the different proteins interact with DNA in different modes.

The *paired* Gene Encodes a Multifunctional Transcription Factor

In addition to an HD, the *paired* gene product, the Prd protein, contains regions that are also found in other developmental genes. We have shown that the so-called Paired domain is a second DNA-binding domain in the Prd protein, which makes Prd a bifunctional transcription factor. Although both the Paired domain and the HD can bind to DNA independently, they can also bind cooperatively to adjacent sites, but only when both domains are present in the same molecule. The cooperation between the two motifs may refine the functional specificity of genes containing highly related domains.

We have undertaken an *in vivo* structure-function analysis of the product of *prd*. In this we are attempting to correlate the multiple DNA-binding functions of Prd defined *in vitro* with the multiple genetic functions of the *prd* gene, as a regulator of segmentation genes and of genes in the nervous system. Transgenic flies that carry the *prd* coding sequence driven by its own promoter rescue the lethal phenotype of prd^- flies. Other lines carrying versions of the *prd* gene modified in regions encoding each of the subdomains of the Prd protein are now being tested for their ability to rescue some of the differential molecular phenotypes of prd^-.

From Segmentation to Organogenesis

We recently discovered a gene that appears to be involved in the early steps of organogenesis of the salivary gland, but not in segmentation. This gene, highly homologous to *prd*, is first expressed in a group of cells that represent the progenitor of the salivary gland placode, at a stage when no tissue differentiation has occurred. This expression is controlled by positional cues from the dorsoventral and anteroposterior determinants. Later, when the placode invaginates, other transcription factor–encoding genes are expressed in the same tissue. Since the expression of this *prd*-related gene precedes morphological events, it may provide an excellent model of transcriptional commitment to a particular differentiation pathway. We are now generating mutants to analyze the early expression of the putative target genes and the resulting morphological phenotype.

Because the protein encoded by this gene has a Prd-like organization but a Prd domain from a divergent class, this molecule will also be very useful in dissecting the residues that determine the specificity of the Prd domain. In collaboration with John Kuriyan's group (HHMI, Rockefeller University), we have undertaken a structural analysis of the Prd domain and the HD present in Prd.

Immune Evasion by Parasites Causing Tropical Diseases

John E. Donelson, Ph.D.—Investigator

Dr. Donelson is also Distinguished Professor in the Department of Biochemistry at the University of Iowa, Iowa City. He obtained his bachelor's degree in biophysics from Iowa State University, served as a Peace Corps volunteer for two years in Ghana, West Africa, and then completed a Ph.D. degree in biochemistry at Cornell University. His postdoctoral research was conducted at the MRC Laboratory of Molecular Biology in Cambridge, England, and at Stanford University. He has received the Iowa Governor's Science Medal and the Burroughs-Wellcome Award in Molecular Parasitology.

THE risk of acquiring parasitic infections is a part of the daily lives of more than 3 billion people living in the developing countries of the tropics. The various protozoan and helminthic parasites that are responsible for these infectious diseases possess a variety of mechanisms for evading the immune response of their hosts. Our laboratory studies the interactions between several of these parasites and the immune system, in anticipation that a more detailed understanding of these immune evasion mechanisms will suggest better ways to combat or prevent the infections. The parasites that cause three of these tropical diseases are described here.

Trypanosomiasis

African trypanosomes are protozoan parasites that cause sleeping sickness or trypanosomiasis throughout equatorial Africa. They are transmitted from tsetse flies to the mammalian bloodstream, where they continually confront the humoral and cellular immune systems. Each trypanosome in the bloodstream is covered by about 10 million copies of a single protein, the variant surface glycoprotein (VSG). The trypanosome population survives the continuous immune assault because individual parasites occasionally switch spontaneously from the expression of one VSG to another—a process called antigenic variation. A new immune response must be mounted against the VSG of the switched parasite and its descendants, enabling the trypanosome population as a whole to stay "one step ahead" of the host immune defenses. We are examining the events at the DNA and RNA level that are responsible for this antigenic variation.

We know that the trypanosome genome contains about 1,000 different genes encoding antigenically distinct VSGs. Usually one, and only one, of these VSG genes is expressed at a time. Rearrangements and duplications of these genes are partly responsible for the selection of which VSG gene is to be transcribed and for the switch event itself.

The rearrangements maneuver specific VSG genes into and out of special chromosomal locations, called expression sites, where transcription occurs. These expression sites are always located near the ends of the chromosomes, i.e., near the chromosomal telomeres, for reasons that are not well understood. The expression process is complicated by the fact that several, and perhaps many, potential telomere-linked expression sites exist in the genome, yet only one is normally activated at any time. Our goal is to understand this activation mechanism at the molecular level.

In one project we have identified a protein that specifically binds to a region upstream of a VSG gene in an expression site activated during the final developmental stage of the parasite in the tsetse fly, i.e., the metacyclic stage. We think that this protein may contribute to the developmental regulation of that VSG gene's transcription. In another project we have detected a high rate of mutation in the telomere-linked VSG genes, which may increase the effectiveness of antigenic variation still further. A third project involves a characterization of a group of very small chromosomes that are unique to African trypanosomes and contain many of the VSG genes destined for sequential expression. In still another project we have introduced into trypanosomes several plasmids containing combinations of sequences upstream of an expressed VSG gene and an easily detected luciferase gene in an effort to identify the sequences that regulate the transcription of the VSG gene.

Leishmaniasis

Leishmania parasites are protozoan pathogens that cause a spectrum of diseases, including cutaneous, mucocutaneous, and visceral leishmaniasis, in many tropical countries. During their life cycle these parasites reside in both the sandfly vector and a mammalian host. Within sandflies they exist as uniflagellar promastigotes that develop from a less-infectious form to the final highly infectious form. This developmental process can be mimicked during cultivation of promastigotes in liquid medium. After transmission from the sandfly to the mammalian host bloodstream, the promastigotes penetrate host macro-

phages, where they reside as spherical amastigotes within acidic phagolysosomes. Thus the parasites evade the immune response by "hiding" inside macrophages—one of several cell types of the immune system that normally help to destroy foreign pathogens and substances. When a sandfly ingests amastigote-laden macrophages during a blood meal, the life cycle is completed. With a grant from the National Institutes of Health, we are studying how these organisms can survive in hostile environments as diverse as a sandfly midgut and an acidic macrophage phagolysosome.

A major glycoprotein on the surface of both promastigotes and amastigotes is a metalloprotease of about 63 kDa (gp63). This enzyme participates in the promastigotes' uptake by macrophages and contributes to the amastigotes' survival within them. The amount of gp63 on the surface of promastigotes increases about 10-fold as the parasites develop into the highly infectious, virulent form during growth in culture. We have discovered that three different RNA species encoding gp63 occur during cultured promastigote growth.

One RNA species occurs in the promastigotes only during their early, logarithmic growth when they are less infectious and have a small amount of gp63 on their surface. This gp63 RNA is replaced by another gp63 RNA species as the promastigotes enter stationary phase and become more infectious. The presence of this second RNA correlates with the increased amount of gp63 protein. The third gp63 RNA species is continuously present at a low level throughout cultured growth and encodes an altered, transmembrane form of the protein.

The three RNA species are derived from a family of about 15 different genes, some of which are identical and all of which are continuously transcribed independent of the stage of cultured growth. The 3′-untranslated regions of the three RNA species have completely different sequences, and we have shown that these sequences contribute to the different steady-state amounts of the RNAs at the different growth stages. In addition, we are introducing recombinant DNAs into leishmania to amplify, alter, or delete individual gp63 genes so that we may determine the functions of each of the three different gp63 proteases during the promastigote and amastigote stages. One of our long-term goals is to alter genetically, or attenuate, the parasite so that it can be readily maintained under laboratory culture conditions but cannot survive in human macrophages. Such parasites may be useful in the development of a vaccine against leishmaniasis.

Onchocerciasis

Onchocerca volvulus is a filarial nematode that causes onchocerciasis, or river blindness, in Africa and Latin America. Female *O. volvulus* worms grow to 50 cm in length and reside throughout the body within nodules. In the nodules they produce thousands of microfilariae each day that migrate throughout the body and enter the eyes, where they produce lesions that can lead to blindness. It is not known how any of the developmental stages of the parasite evade the immune response of an infected person. The parasites are difficult to study in the laboratory, because there is no good experimental animal model; they infect only humans and chimpanzees.

We have constructed cDNA expression libraries of the mRNAs from the infective L3 stage larvae of the parasite, and we are using specific cDNA clones to overproduce parasite antigens from this larval stage that are recognized by antisera from onchocerciasis patients. This approach has revealed several proteins that are unique to the parasite and may be valuable for improved diagnosis, treatment, and prevention of the disease.

Post-transcriptional Regulation of Gene Expression, RNA-Protein Complexes, and Nuclear Structures

Gideon Dreyfuss, Ph.D.—Investigator

Dr. Dreyfuss is also Professor of Biochemistry and Biophysics at the University of Pennsylvania School of Medicine. He received his Ph.D. degree in biological chemistry from Harvard University and his postdoctoral research training as a Helen Hay Whitney fellow at the Massachusetts Institute of Technology. Prior to his present appointment, he was Professor and Established Investigator of the American Heart Association at Northwestern University.

MESSENGER RNAs (mRNAs) are the functional translatable intermediates in the pathway of gene expression from DNA to proteins. They are formed in the nuclei of eukaryotic cells by extensive and tightly regulated post-transcriptional processing of primary RNA polymerase II transcripts. These transcripts are termed heterogeneous nuclear RNAs (hnRNAs), a term that describes their size heterogeneity and cellular localization. The terms hnRNA and pre-mRNA are often used interchangeably, though it is possible that only a subset of hnRNAs are actually precursors to mRNA and that the rest turn over in the nucleus. From the time hnRNAs emerge from the transcription complex and as long as they remain in the nucleus, they are associated with proteins.

The collective term for the proteins that bind hnRNAs (but are not stable components of other classes of ribonucleoprotein complexes, such as snRNPs [small nuclear ribonucleoproteins]) is hnRNPs. The significance of hnRNPs is that they are bound to the hnRNAs and thus influence their structure and therefore their fate and processing into mRNAs. They are also abundant in the nucleus and in hnRNA-hnRNP complexes (hnRNP complexes) and of interest as major nuclear structures.

Once formed, the mRNAs are transported to the cytoplasm via nuclear pores—a process that appears to be one of the most important regulatory steps in the post-transcriptional pathway of gene expression and about which very little is presently known. In the cytoplasm, mRNAs are associated with proteins—the mRNPs—and these are likely to be involved in the regulation of the translation and stability of mRNAs and in their cellular localization. Our goal is to understand, in molecular detail and cellular architecture, how the post-transcriptional portion of the pathway of gene expression operates in the cell. To that end, we investigate the structure, function, and localization of the hnRNPs and mRNPs and the RNP complexes.

We have used photochemical RNA-protein crosslinking in intact cells and affinity chromatographic methods to identify and purify the hnRNPs and mRNPs and have produced monoclonal antibodies to many of them. The monoclonal antibodies were used for immunopurification of hnRNP complexes from vertebrate and *Drosophila melanogaster* cells and for the characterization of the hnRNPs. Immunopurified hnRNP complexes contain large RNA of up to 10,000 nucleotides and at least 20 major proteins, designated A–U, in the range of 34–120 kDa. There are also many less abundant hnRNPs, and these appear to bind only to specific subsets of hnRNAs. The specific arrangement of the proteins on specific hnRNAs is probably important in determining the structure of the hnRNA and is one of the areas on which we concentrate our investigation.

Related to this issue, we found that several of the hnRNPs have RNA-binding specificities. Interestingly, some of the specificities of the major hnRNPs are for sequences important in pre-mRNA processing and polyadenylation, and it is likely that this binding specificity is directly related to a role for these proteins in mRNA formation.

The molecular cloning and sequencing of cDNAs for several RNPs made possible the discovery of a conserved RNA-binding domain (RBD) and a ribonucleoprotein consensus sequence (RNP-CS). The RNP-CS, Lys/Arg-Gly-Phe/Tyr-Gly/Ala-Phe-Val–X–Phe/Tyr, is the most highly conserved segment in a generally conserved domain of about 90 amino acids found in many RNA-binding proteins of the nucleus and cytoplasm in all eukaryotes examined. Overall, RNP-CS proteins have a modular structure reminiscent of DNA-binding transcription factors. They frequently contain several similar but nonidentical RBDs, and all contain at least one auxiliary domain that is unique to each type of protein (e.g., glycine-rich, glutamine-rich, or acidic) and that most likely functions in protein-protein interactions.

The hnRNP C1/C2 proteins are abundant, avid hnRNA binders. Deletional analysis showed that a 93–amino acid segment of the human hnRNP C protein that contains the RNP-CS is, as predicted,

sufficient for RNA binding. This RBD was produced in bacteria and purified to homogeneity in active form. Nuclear magnetic resonance (NMR) methods (in collaboration with Luciano Mueller and Michael Wittekind, Bristol-Myers Squibb), including ^{13}C- and ^{15}N-edited three-dimensional NMR, were used to determine the structure of the RBD in solution. The compact folded structure exhibits a four-stranded antiparallel β-sheet and two well defined α-helices. The structure of this RBD complexed with an RNA substrate is being determined.

Experiments on mitotic cells unexpectedly provided important insights into the assembly and general nature of hnRNP complexes and into the transport of proteins to the nucleus. In mitosis, as the nuclear envelope breaks down, hnRNPs disperse throughout the cell but remain associated in complexes with RNA. After mitosis, once the nuclear envelope re-forms, preexisting hnRNPs return to the nucleus. We observed, however, using double-label immunofluorescence microscopy with monoclonal antibodies to various hnRNPs on postmitotic cells, that at the end of mitosis the hnRNP complexes dissociate in the cytoplasm and the different proteins are transported to the nucleus separately. Some, including C1, C2, and U, like snRNPs and lamins, are transported immediately (early group), while others, including A1, A2, B1, B2, E, G, H, and L, are transported into the nucleus later (late group). Thus, immediately following reassembly of the nuclear envelope at the end of mitosis, pairs of cells are detected in which some hnRNPs are in the nucleus, and others are in the cytoplasm. These observations show that hnRNP complexes are dynamic structures, in that hnRNPs can dissociate from the complexes and return to the nucleus separately.

Surprisingly, the transport of the late group requires transcription by RNA polymerase II: inhibitors of this polymerase cause the late proteins to remain in the cytoplasm. Thus there are two pathways for nuclear transport localization of proteins: a transcription-independent pathway and a novel transcription-dependent pathway. The different hnRNPs utilize one of the two pathways, and both pathways operate throughout the cell cycle. The signals in the proteins that specify the pathway to be used, the mechanism of transcription-dependent transport localization, and the relevance to mRNA transport are being investigated.

Immunofluorescence microscopy indicated that the hnRNP proteins A, B, C, E, I, K, L, M, and U are nucleoplasmic—that is, localized to the nucleus but excluded from nucleoli. In contrast, the mRNPs that have been characterized so far [in particular the poly(A)-binding protein] are confined to the cytoplasm. It was therefore concluded that the mRNA must exchange most if not all of the proteins with which it is associated in the nucleus as it is transported to the cytoplasm. The dissociation of these proteins from the mRNA and the subsequent binding of mRNPs must be an important aspect of nuclear-cytoplasmic transport of mRNA.

Recently we found that several of the abundant hnRNPs, including A1, are not confined to the nucleus but rather shuttle continuously between the nucleus and the cytoplasm. Thus hnRNPs may have cytoplasmic functions in addition to their nuclear roles in the processing of pre-mRNA to mRNA. Furthermore, A1 is bound to mRNA in the cytoplasm and its return to the nucleus requires RNA polymerase II transcription. It is therefore possible that the novel cytoplasmic ribonucleoprotein complex of mRNA with hnRNPs is the substrate of RNA nuclear-cytoplasmic transport.

Thus it is likely that hnRNPs play an active role in mRNA export and that the mRNA is transported to the cytoplasm as a result of its association with the shuttling hnRNPs. Understanding the trafficking of hnRNPs in the cell and the mechanisms that regulate the assembly and disassembly of RNPs with hnRNAs and mRNAs should be important for elucidating the post-transcriptional pathway of gene expression, including the nuclear transport process of mRNA.

Genetic Basis of Hearing Loss

Geoffrey M. Duyk, M.D., Ph.D.—Assistant Investigator

Dr. Duyk is also Assistant Professor of Genetics at Harvard Medical School and a member of the Eaton-Peabody Laboratory at the Massachusetts Eye and Ear Infirmary. He received his undergraduate degree at Wesleyan University and his M.D. degree and a Ph.D. degree in biochemistry at Case Western Reserve University. After internship and residency at the University of California, San Francisco, he held a fellowship in medical genetics under the direction of Charles Epstein. His postdoctoral research work was also at UCSF, with David Cox and Richard Myers. While at UCSF, Dr. Duyk was awarded an HHMI Physician Research Fellowship and was a Lucille P. Markey Scholar in Biomedical Science.

HEARING loss is the commonest form of sensory impairment. In the United States the incidence of congenital deafness is approximately 1 per 1,000, and at least half of these cases are likely to be determined genetically. It is estimated that by age 65 one in six of us will have a clinically significant hearing loss. While environmental factors play a role in hearing impairment, genetic factors often determine predisposition. Deafness is a major aspect of over 100 defined genetic disorders.

Our laboratory is interested in sensorineural hearing loss in which the abnormality should lie along the pathway between the sensory receptors of the inner ear and the auditory centers of the brain. Our strategy is to combine genetic mapping and positional cloning with candidate gene approaches to identify genes responsible for hearing loss and to further our understanding of the molecular mechanisms of hearing.

Candidate Genes

Within the inner ear reside the sensory organs for hearing and balance. Both of these systems represent mechanical transduction. In each, a specialized "hair cell" serves as a sensor of sound or motion. Hair cells are highly polarized, with the basal surface forming synapses with afferent and efferent nerve fibers. Finger-like projections referred to collectively as the hair cell bundle are distinctive features of the hair cell's apical surface. Deflection of the hair cell bundle by the sensory input activates mechanically sensitive ion channels, initiating signal transduction. This process is distinct from other sensory systems such as taste, olfaction, and vision, where a ligand binds to a specific receptor, which in turn activates a second messenger system to initiate sensory transduction. The molecular components of this highly specialized mechanical sensory system are logical targets for specific mutations resulting in hearing loss.

In collaboration with David Corey (HHMI, Massachusetts General Hospital), we have begun the process of identifying the molecular components of the hearing apparatus by combining molecular biological and biophysical approaches. Although extensive microanatomical and biophysical characterization of this apparatus has been undertaken, little biochemical or molecular information is available. As a starting point, we are constructing cDNA libraries from microdissected inner ear material highly enriched for hair cells. In addition, we are exploring strategies for analyzing mRNA populations of single cells with the polymerase chain reaction (PCR) technique to augment the traditional approaches for recovering and studying genes from specialized tissue sources. As we identify components of the sensory transduction pathway, human and murine homologues will be recovered and mapped to specific chromosomes, and closely linked DNA polymorphic markers will be identified. The derived information will be an important resource as we begin our search for the mutations responsible for this group of inherited disorders.

"Reverse Genetics"

The genetic analysis of hearing loss in human populations is complicated by the difficulty of differentiating many of these syndromes (genetic heterogeneity) and the fact that deaf individuals often intermarry (nonassortive mating). As a consequence, most of the families suitable for study are small, increasing the problem of placing the disease locus with the degree of precision required to clone the gene. The availability of mapped candidate genes helps to bridge the gap between genetic and physical mapping. Analysis of these small families will be enhanced by the availability of dense genetic maps. Toward that end, we have developed an efficient technique for the construction of genomic libraries augmented for selected classes of simple sequence repeats (SSRs). SSRs correspond to runs of di-, tri-, or tetranucleotide sequences (e.g., $[CA]_n$ or $[GGC]_n$) that often demonstrate length polymorphisms detectable by PCR assay. These libraries will aid in the production of high-resolution ge-

netic maps composed of very polymorphic, homogeneously distributed PCR-based markers.

Genetic mapping can localize a disease gene to a 1- to 2-Mb region of a particular chromosome. The next challenge is to identify the genes in this region. We have developed a strategy, termed exon trapping, that utilizes splicing signals as primary identifiers of coding sequence in cloned genomic DNA. This approach is also of great utility for cloning the human homologue of a gene of interest when the temporal or spatial pattern of gene expression limits the availability of mRNA for analysis. We are developing second-generation exon-trapping strategies based on large-capacity cloning vectors (e.g., cosmids, P1) that should extend the range for gene searching and help to integrate the recovery of genes with physical mapping.

Molecular Genetics of Intracellular Protein Sorting

Scott D. Emr, Ph.D.—Associate Investigator

Dr. Emr is also Professor of Cellular and Molecular Medicine at the University of California, San Diego, School of Medicine. He did his graduate work in microbiology and molecular genetics with Thomas Silhavy and Jonathan Beckwith at Harvard Medical School, where he identified and characterized the first secretion-defective signal sequence mutants as well as the first component of the bacterial protein export apparatus. Dr. Emr did postdoctoral work on protein secretion in yeast with Randy Schekman at the University of California, Berkeley. He was a faculty member at the California Institute of Technology before moving to UCSD. Dr. Emr counts among his honors a Searle Scholars Award and an NSF Presidential Young Investigator Award.

An essential feature of all eukaryotic cells is their highly compartmentalized organization. Different, often competing biochemical processes are segregated into distinct compartments. The identity, stability, and function of each of these compartments, or organelles, is conferred in large part by the unique set of proteins that reside within it. These proteins must be routed from their common site of synthesis in the cytoplasm to their unique site of function in the appropriate intracellular organelle. Toward a detailed understanding of the molecular mechanisms that direct the delivery of one class of these proteins, we have focused our attention on the transport and sorting of proteins through the Golgi complex to the lysosome.

Lysosomal Hydrolase Sorting

Our laboratory is using a simple unicellular eukaryote, the yeast *Saccharomyces cerevisiae*, as a model genetic system to study protein trafficking through the secretory pathway to the lysosome-like vacuole. The fundamental similarities between yeast and other eukaryotic cells in their pathways and mechanisms for protein delivery have clearly established yeast as an important model system for the study of these problems.

The importance of the lysosomal protein sorting pathway in humans is revealed when the effects of mislocalizing such enzymes are examined. A number of lysosomal hydrolases are secreted from naturally occurring tumor cells (e.g., cathepsin D by breast cancer cells). It has been proposed that the mislocalization of lysosomal enzymes enhances cell growth and increases the metastatic potential of tumor cells by contributing to the hydrolysis of extracellular matrix components of target tissues. The importance of this sorting pathway is further exemplified by the serious inherited disorders (e.g., I cell disease) that result in mislocalization of lysosomal hydrolases.

Proteins destined for the vacuole/lysosome in yeast and mammalian cells transit through early stages of the secretory system. The transport and processing characteristics of several vacuolar hydrolases in yeast, including the soluble protease carboxypeptidase Y (CPY), have been well characterized. CPY is translocated into the endoplasmic reticulum (ER), where it is modified with four core oligosaccharides to generate the ER precursor form of CPY (p1CPY). Subsequent transport events are mediated by vesicular carriers. Delivery from the ER to the Golgi complex and passage through the Golgi are accompanied by elongation of the core oligosaccharides on CPY, resulting in formation of the Golgi-modified form of CPY (p2CPY). Sorting of p2CPY from proteins destined for secretion appears to take place in the late Golgi. Upon arrival in the vacuole, an amino-terminal propeptide segment on p2CPY is proteolytically removed, generating the active mature form of the protease (mCPY).

Sorting-Defective Mutants

We have focused our efforts on identifying and characterizing the cellular machinery that directs the sorting and transport of vacuolar hydrolases. A gene fusion-based selection scheme we designed enabled us to isolate more than 600 yeast mutants that exhibit severe defects in vacuolar protein sorting. Each of the mutants missorts and secretes CPY and other vacuolar enzymes. The recessive mutations define more than 33 complementation groups referred to as *vps* (vacuolar protein sorting defective). Extensive genetic, biochemical, and morphological characterization of the *vps* mutants has allowed us to organize them into phenotypically related groups. Each group appears to function at a common stage of the protein sorting pathway.

The apparent genetic complexity of the *vps* mutant collection presumably reflects the biochemical complexity of the protein sorting reaction. Specific cellular components must recognize vacuolar proteins, segregate them from other secretory proteins, and package them into

specific transport vesicles that ultimately must recognize and fuse with the appropriate target organelle, the vacuole. Therefore the list of potential activities and structures required for the sorting and transport of lysosomal/vacuolar proteins can easily accommodate the large number of gene products presently implicated by the genetic studies.

Role for a Protein Kinase Complex

An understanding of this protein sorting pathway and the individual activities of different *VPS* gene products is being facilitated by the molecular isolation and characterization of these genes. Thus far, we have cloned and sequenced eight *VPS* genes. Comparison of the yeast *VPS* gene sequences with other known mammalian genes has revealed several informative structural and functional similarities. Two of the genes, *VPS*15 and *VPS*34, are of particular interest. The sequence of the *VPS*15 gene predicts a protein with the following features: a consensus site for amino-terminal myristoylation, a region that shares significant sequence similarity with the family of Ser/Thr protein kinases, and a region of homology with the regulatory subunit of protein phosphatase 2A (PP2A). The sequence similarity Vps15 protein (Vps15p) shares with Ser/Thr protein kinases and the PP2A regulatory subunit raises the interesting possibility that protein phosphorylation/dephosphorylation may play an important role in the regulation of protein sorting events.

To assess the functional significance of these sequence similarities, we used site-directed mutagenesis to change several highly conserved amino acid residues in the putative kinase domain of the Vps15p. Each of the mutations inactivates the complementing activity of the *VPS*15 gene. The mutant cells exhibit extreme vacuolar protein sorting defects; amino acid changes in less highly conserved positions gave only weak protein sorting defects.

These data suggest that the Vps15p is an active kinase and that this kinase activity is required during some step in vacuolar protein sorting. Protein phosphorylation may act as a "molecular switch" in this protein sorting pathway by actively diverting vacuolar hydrolases away from the default secretion path and toward the vacuole. One can imagine several points in the pathway that may need to be regulated precisely, such as the budding and transport of carrier vesicles or the recognition and fusion of these with the correct target organelle.

Biochemical and genetic evidence indicate that the Vps15 protein forms a complex (on the cytoplasmic face of the membrane) with another Vps protein, Vps34p. Overexpression of Vps34p suppresses the vacuolar protein sorting and growth defects caused by mutations within the kinase domain of Vps15p, but will not suppress a null allele of *VPS*15. Therefore, Vps34p cannot bypass the cells' requirement for Vps15p. This genetic interaction between *VPS*15 and *VPS*34 is consistent with the observation that mutations in both the *VPS*15 and *VPS*34 genes result in a common set of phenotypes.

The Vps34 protein shares sequence similarity with a mammalian gene recently identified in Mike Waterfield's laboratory that codes for the catalytic subunit of the phosphatidylinositol 3-kinase (PI3-kinase). PI3-kinase catalyzes the formation of PI3-phosphate, a rare membrane lipid that has been proposed to act as a second messenger in cell signaling. The enzyme appears to play an important role in cell proliferation and transformation. PI3-kinase activity has been shown to be associated with several cell surface protein-tyrosine kinase receptors (e.g., the PDGF [platelet-derived growth factor], insulin, and CSF-1 [colony-stimulating factor 1] receptors). The role of this lipid-modifying enzyme in protein trafficking events is not yet clear. However, we are in the process of mutating several conserved sequence motifs to analyze the significance of this intriguing sequence similarity in Vps34 protein function.

Among its functions, PI3-phosphate formation at localized sites in the membrane (adjacent to membrane receptor proteins) may facilitate vesicle formation/targeting required during protein transport to the lysosome from both the Golgi complex (delivery of newly synthesized lysosomal hydrolases) and the cell surface (endocytic uptake and down-regulation of cell surface receptors). The Vps15 kinase may regulate the activity of Vps34p/PI3-kinase by phosphorylation of Vps34p.

We recently developed an *in vitro* assay that reconstitutes intercompartmental protein transport to the yeast vacuole. We are now using this assay to assign the Vps15 and Vps34 proteins to a specific step(s) in the reaction.

Mechanisms Involved in Preventing Unwanted Blood Clots

Charles T. Esmon, Ph.D.—Investigator

Dr. Esmon is also a member of the Oklahoma Medical Research Foundation and Professor of Pathology and Associate Professor of Biochemistry at the University of Oklahoma Health Sciences Center, Oklahoma City. He received his B.S. degree in chemistry from the University of Illinois and his Ph.D. degree in biochemistry from Washington University. He conducted his postdoctoral research at the University of Wisconsin before joining the faculty at the University of Oklahoma Health Sciences Center. Later he joined the Oklahoma Medical Research Foundation.

PROTEIN C, protein S, and thrombomodulin constitute one of the natural anticoagulant complexes that prevents unwanted blood clots. We have focused our attention on this system because we have been able to identify patients with a history of unwanted blood clots who have abnormal protein C, protein S, or thrombomodulin. To understand how the system functions, it is useful to review the function of the components. Thrombomodulin is found primarily on the surface of endothelial cells, the cells that line the blood vessels. Thrombin, the enzyme that causes blood to clot, can bind to thrombomodulin; when thrombin is bound, it no longer clots the blood but instead converts protein C into the active blood clotting inhibitor, activated protein C. Activated protein C then binds to protein S on the surface of platelets (small cells in the blood) or endothelial cells, where it functions as an anticoagulant. The activated protein C–protein S complex works as an anticoagulant by cutting up and inactivating two of the clotting proteins, factor VIII (the protein missing in hemophilia) and factor V.

This broad outline of how the system functions fails to tell us much about where, when, or how the system might function in human disease processes. This knowledge is important both in terms of understanding the basic properties of the system and in the design of new therapeutic approaches to diagnosis and prevention of blood clots. Of particular interest, and still unexplained, is the observation that administration of activated protein C at levels that can prevent unwanted blood clots does not increase blood loss at surgical sites. This contrasts with available anticoagulants, such as heparin, which block unwanted clot formation but also dramatically increase blood loss at surgical sites. One of our goals is to understand how this natural anticoagulant can accomplish this remarkable specificity. New therapeutic agents with these properties could greatly decrease morbidity and mortality associated with thrombotic complications.

Most healthy individuals have adequate amounts of protein C and the other components of the system that control blood clot formation under normal circumstances. When people become sick, unwanted clotting is often a problem. Studies from other laboratories have shown that protein S circulates in humans both free and bound to an inhibitor of the complement system (the system that helps protect from infection), called C4b-binding protein (C4bBP).

We found that only the free form of protein S could work to form the anticoagulant. Patients with clinical conditions known to cause an increased risk of blood clots also had reduced levels of free protein S and more C4bBP–protein S complex. Families with inherited thrombotic complications were identified in which the family members who developed blood clots had high levels of the complex. These observations suggested that alteration in the levels of free protein S might contribute to the clotting complications observed in these patients. To test this hypothesis, we developed some animal models that involved blood clotting as a complication.

Clearly thrombosis is caused by more than simply an alteration of C4bBP levels. Infection and agents that cause inflammation are known to trigger coagulation by a variety of mechanisms. Three very different responses occur. The first is related to bacterial infection, which can result in septic shock, a process characterized by small blood clots in the circulation, damage to organs, and death. Once this process begins, treatment is very difficult. A second response is formation of solid blood clots in the small vessels. This results in death of the affected organs. A third response is occlusion of the large vessels. Why inflammation causes these three dramatically different responses is unknown, complicating rational approaches for both early diagnosis and effective therapy.

Animal models for these three processes were developed in collaboration with Fletcher Taylor. These models indicate that shifts in the balance between free protein S and protein S bound to C4bBP contribute to the different disease processes and can be involved in shifting the observed type of thrombosis. Increased levels of

C4bBP–protein S complex cause the inflammatory response to low levels of bacteria to convert from a mild response to a severe response with all of the characteristics of toxic shock. Moreover, under certain conditions the shift in amounts of C4bBP–protein S complex can change the response from that of circulating small clots to occlusion of small vessels. Most data suggest that bacteria cause clotting by causing the formation of inflammatory mediators, called cytokines. Inflammatory cytokines do not lead to the small clots characteristic of bacterial infection but rather to blood clots in the major vessels. This response occurs only when the C4bBP–protein S levels are high. In every case tested to date, the deleterious effects of the C4bBP can be reversed by adding protein S.

Why then do bacterial infections lead to the small clots in the circulation? Possible insights were gained from the observation that small cell fragment-like particles can convert this thrombotic response back into the small-clot syndrome seen in septic shock. Complement is known to cause such cell fragments to be formed. These studies may provide insights into how different inflammatory and coagulation components work in concert to generate such distinct—but related—thrombotic complications.

These results suggest that the balance between free protein S and that bound to C4bBP may be critical to thrombotic complications as a result of inflammation. The data imply that increasing protein S would prevent these thrombotic complications, suggesting new therapeutic approaches to thrombotic diseases. If we could block the decrease in free protein S, the risk of thrombosis might be reduced. This unique approach could potentially return the patient to normal status without significantly increasing the risk of bleeding.

A major interest in our laboratory has been to understand how thrombomodulin causes thrombin to change its function from a clotting to a clot-inhibiting enzyme. To understand this, we have examined the binding of thrombin to thrombomodulin and to its other targets. Of particular interest is the ability of thrombin to activate cells and platelets, leading to platelet plugs and clots, especially in the arteries. Thrombomodulin can block platelet activation by thrombin, but how it does so had been a mystery. In collaboration with Shaun Coughlin's group, we demonstrated that the thrombin receptor on platelets and thrombomodulin bind thrombin at the same place, thereby explaining the earlier observations. Both protein C and the thrombin receptor share common sequences near the cleavage sites, and both sequences are inherently poor sequences for cleavage by thrombin. Binding of thrombomodulin fragments or the fragments from the thrombin receptor results in changing how thrombin cleaves its substrates. Thus both receptors have a built-in switch that allows thrombin to function.

Molecular Genetics of Steroid and Thyroid Hormone Receptors

Ronald M. Evans, Ph.D.—Investigator

Dr. Evans is also Professor at the Gene Expression Laboratory of the Salk Institute for Biological Studies and Adjunct Professor in the Departments of Biology and of Biomedical Sciences at the University of California, San Diego. He received his Ph.D. degree in microbiology and immunology from the University of California, Los Angeles, School of Medicine. After postdoctoral training with James Darnell at the Rockefeller University, he joined the faculty of the Salk Institute. Dr. Evans is a member of the National Academy of Sciences. His research interests are in developmental and inducible regulation of gene expression.

An understanding of the mechanisms by which apparently distinct regulatory systems integrate to modulate body function and behavior poses one of the most important challenges of modern biology. Hence we have focused our attention on the action of steroid, retinoid, and thyroid hormones in regulatory cellular and organ physiology. This field has undergone an extraordinary development in the last several years as a consequence of the cloning and sequencing of the genes encoding the receptors for these hormones in target cells.

It has been demonstrated that these receptors are all structurally related and constitute a superfamily of nuclear regulatory proteins that are capable of modulating gene expression in a ligand-dependent fashion. One challenge is to define each receptor's molecular properties that determine its interactions with the transcription machinery regulating gene expression. Another challenge is to elucidate the contributions of individual regulatory systems to the integrated and complex processes associated with cell growth, differentiation, and organ function.

A Novel Retinoic Acid Response Pathway

The retinoids, a group of compounds that include retinoic acid, retinol (vitamin A), and a series of natural and synthetic derivatives, exert profound effects on development and differentiation in a wide variety of systems. (The retinoic acid receptors are collectively designated RARs.) Retinoic acid has also been shown to induce the transcription of several genes, suggesting a role analogous to those of steroid and thyroid hormones. In previous studies we described the cloning and characterization of a retinoic acid–dependent transcription factor referred to as RARα. Additional RAR-related genes have been isolated, and at least three different RAR subtypes (α, β, and γ) are now known in mice and humans.

Retinoic acid receptors share homology with the superfamily of steroid and thyroid hormone receptors and have been shown to regulate specific gene expression by a similar ligand-dependent mechanism. Complicating these observations is our recent identification of a group of receptors termed RXRs (retinoid X receptors), which are only distantly related to the RARs. The discovery of this second retinoid transduction pathway led us to investigate its functional properties and determine its relationship to the RARs. We now know that there are at least three RXR-related genes (termed α, β, and γ) located at genetically distinct loci. Northern blot analyses of the RXRs indicate that each isoform has a unique pattern of expression in adult tissue and is temporally and spatially expressed in the embryo. These studies suggest a role for RXRs in adult physiology and embryonic development.

Binding experiments demonstrate that the RXR protein has low affinity for retinoic acid (RA) and, taken together with the transactivation studies, indicate that the RXR ligand may be a metabolite of RA. Based on these assumptions, we devised a strategy to identify the putative metabolite referred to as retinoid X. The implicit concept of the strategy was that all-*trans* RA may be converted to retinoid X by a natural cellular process. Accordingly, high doses of all-*trans* RA were fed to recipient tissue culture cells. After allowing for metabolic conversion, material extracted from these cells was fractionated to resolve the various retinoid peaks, and each peak was assayed for its ability to activate the RXRs in a transfection assay. A specific peak, termed retinoid X, was identified and characterized by mass spectrometry.

Based on this process, we have now demonstrated that 9-*cis* RA is the high-affinity ligand for the RXR. While not previously seen in living organisms, 9-*cis* RA is apparently a new and widely used vertebrate hormone. It transactivates RXRα up to 40 times more efficiently than all-*trans* RA and binds to RXR with high affinity. We also confirmed that all-*trans* RA shows no detectable binding affinity for the RXR. Furthermore, each of the RXR subtypes (α, β, and γ) is activated by 9-*cis* RA with increased potency and efficacy relative to all-*trans* RA.

A point of potential physiological significance is that 9-*cis* RA also binds to and transactivates both RXRs and RARs and may thus serve as a common or "bifunctional" ligand. Conversion of the all-*trans* to the 9-*cis* isomer could provide a novel means for differential cell-specific regulation of the activity of these retinoid pathways. The hypothesis that 9-*cis* RA may be functionally distinct from its all-*trans* precursor raises the intriguing possibility that the regulation of its isomerization could be a key step in retinoid physiology. It is unknown whether this reaction is catalyzed by an enzyme.

The 3-4-5 Rule

Members of the receptor superfamily modulate target gene expression by binding as either homo- or heterodimers to hormone response elements (HREs). We recently described the properties of direct repeats of the consensus half-site sequence AGGTCA as HREs for nuclear receptors. Receptor specificity for binding and activation was shown to be conferred through the number of nucleotides separating the two half-sites. Spacers of 3, 4, or 5 nucleotides were originally shown to serve as optimal response elements for the vitamin D receptor (VDR), thyroid hormone receptor (TR), and RAR, respectively. We now refer to this physiological code built into HREs as the "3-4-5 rule."

More recently, we have characterized in the upstream regulatory region of the cellular retinol-binding protein type II (CRBPII) gene an HRE that confers selective responsiveness to the RXRs. This response element consists of tandem repeats of the AGGTCA sites separated by a single nucleotide.

The ability of the VDR, TR, and RAR to recognize their cognate response elements is dependent upon their ability to form a heterodimeric complex with an unknown nuclear factor. Remarkably, the search for this factor has recently identified RXR as a common heterodimeric partner for the VDR, TR, and RAR. These results emerged from our initial finding of a functional interaction between the RAR and RXR, followed by the demonstration that these two receptors form stable heterodimers in solution. Furthermore, the heterodimer binds target DNA with more than 100-fold increased efficiency over either partner alone.

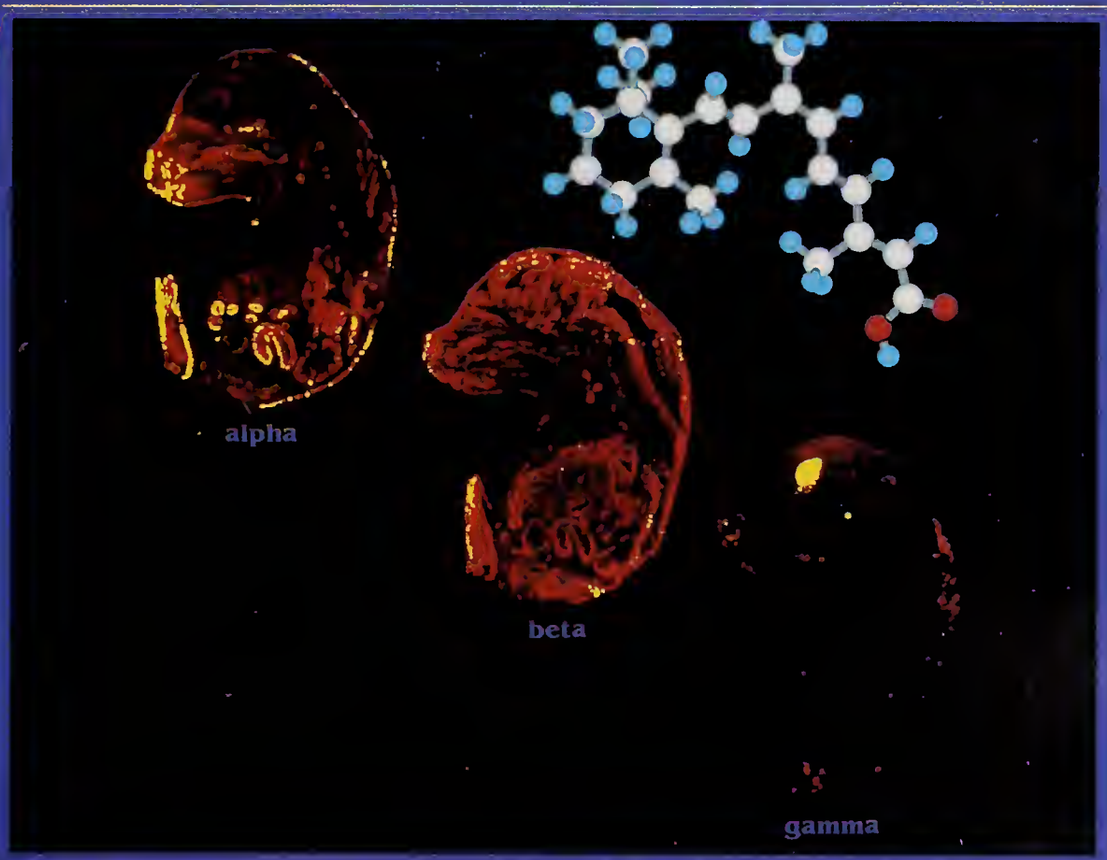
Similarly, RXR heterodimers with the VDR and TR have greatly increased affinity for the respective cognate response elements. Apparently RXR is serving as a type of master receptor, gating the activities of vitamin D, thyroid hormone, and retinoic acid. As noted above, however, RXR can act independently of all these receptors when responding to its novel hormone 9-*cis* RA. Despite the apparent level of complexity revealed by these interactions, it appears that the physiological response is built upon a series of simple principles in which heterodimers, each consisting of an RXR component, interact with common DNA sequences, each varying by a single nucleotide. It is through this simplicity that such exquisite specificity can be maintained and thus permit coordinate control of a vast gene network in a highly selective and orderly fashion.

*Opposite: Expression pattern of the genes for RXRα, β, and γ in a mouse embryo. RXRs are receptor proteins that mediate the actions of the newly described hormone, 9-*cis *retinoic acid. Parasagittal sections from the embryo taken at gestation day 16.5 were hybridized* in situ *with radioactive probes that mark the specific areas where RXR mRNA is expressed. After exposure to x-ray film, the resulting image is digitally scanned into a computer and colorized. Dark red to yellow colors represent low to high levels of RXR expression. Thus the figure demonstrates that RXRα is strongly expressed in skin and metabolic organs such as intestine; RXRβ is expressed ubiquitously; RXRγ shows marked expression in the pituitary and corpus striatum, the brain center that regulates muscle coordination and is affected in parkinsonian disorders.*

Reprinted with permission from Mangelsdorf, D.J., Borgmeyer, U., Heyman, R.A., Zhou, J.Y., Ong, E.S., Oro, A.E., Kakizuka, A., and Evans, R.M. 1992. Genes Dev *6:329–344. Computer photograph by Jamie Simon, the Salk Institute for Biological Studies.*

GENES & DEVELOPMENT

VOLUME 6 NUMBER 3 March 1992



Cold Spring Harbor Laboratory Press
in association with
The Genetical Society of Great Britain

Molecular Mechanisms Involved in the Actions of Calcium-mediated Hormones

John H. Exton, M.D., Ph.D.—Investigator

Dr. Exton is also Professor of Molecular Physiology and Biophysics and of Pharmacology at the Vanderbilt University School of Medicine. He received his medical degree from the University of New Zealand and his Ph.D. degree in biochemistry from the University of Otago, New Zealand. His postdoctoral research was done in the Department of Physiology at Vanderbilt University School of Medicine with Charles Park, where he has remained as a faculty member. His honors include the Lilly Award of the American Diabetes Association and the M.D. degree with distinction from the University of Otago.

THE major objective of my laboratory is to elucidate the mechanisms of action of hormones, neurotransmitters, and other agents that transmit information in the nervous system and other organ systems by altering membrane lipids and increasing the concentration of calcium ions in their target cells. Agents that act this way include regulators of heart function and blood flow, such as epinephrine, norepinephrine, acetylcholine, angiotensin, and vasopressin; other neurotransmitters, such as serotonin, neurotensin, and substance P; and agents that control certain pituitary and pancreatic secretions, food digestion, bladder and uterine contraction, platelet aggregation, and certain responses to trauma and infection.

We initially established that many hormones and neurotransmitters act by increasing the intracellular concentration of calcium ions. The next phase of our work involved the demonstration that the increase is due to both mobilization of calcium ions from internal stores and stimulation of their inflow across the cell membrane. We also demonstrated that the receptors for calcium-mediated agents are located on the outer surface of their target cells. Thus these cells must have some means of signaling from the receptors to the internal calcium stores, and our efforts were directed toward elucidating the signaling mechanism.

Initially we tested the hypothesis that the signal is generated by the breakdown of phosphatidylinositol, a phospholipid in the cell membrane. However, this was found to be too slow to account for the changes in calcium. Attention was then focused on a related phospholipid, phosphatidylinositol 4,5-bisphosphate (PIP_2), which breaks down more rapidly in response to hormones. The situation became clearer when inositol 1,4,5-trisphosphate (IP_3) was identified as the signaling molecule for intracellular calcium release. This compound is generated when PIP_2 is broken down by the enzyme phospholipase C. The other compound produced is 1,2-diacylglycerol (DAG), which is also a signaling molecule, since it activates a specific protein-phosphorylating enzyme, protein kinase C.

The present activities of the laboratory encompass two major research areas. The first involves elucidating how calcium-mediated agents stimulate the breakdown of PIP_2. A major discovery has been the finding that a G protein (a regulatory protein that binds the nucleotide GTP) is involved in coupling the receptors for these agents to the phospholipase C enzyme that breaks down PIP_2. Our work has involved purifying and characterizing the relevant G proteins from liver cell membranes and reconstituting them with other components of the signaling system. Reconstitution of the G proteins with PIP_2 phospholipase C has been achieved, and the system has been used to purify the G proteins to homogeneity in both the complete form ($\alpha\beta\gamma$ heterotrimers) and in the form of free α-subunits. Two α-subunits (42 and 43 kDa) have been identified, and both have been shown immunologically to be members of the G_q family of G proteins. Partial sequencing of tryptic peptides of both proteins has confirmed that they are $G\alpha_q$ and $G\alpha_{11}$.

Purification of the G protein activators of the phospholipase C in the heterotrimeric form has likewise yielded two G proteins corresponding to G_q and G_{11}. Using $G\alpha_q$ and $G\alpha_{11}$ activated by a GTP analogue, we have shown that the specific isozyme form of the PIP_2 phospholipase C controlled by these G proteins in both liver and brain is the 148-kDa β_1-isozyme. Evidence that G_q and G_{11} are coupled to receptors for calcium-mobilizing agents has also been obtained. Three calcium-mediated agents (vasopressin, epinephrine, and angiotensin) specifically stimulate the labeling of $G\alpha_q$ and $G\alpha_{11}$ by a radioactive, light-reactive analogue of GTP in liver cell membranes. Binding of a GTP analogue is enhanced by an acetylcholine analogue when G_q and G_{11} are reconstituted with the M_1 (calcium-mobilizing) muscarinic receptor but not with the M_2 receptor. Co-reconstitution of the purified M_1 receptor, G_q/G_{11}, and phospholipase C-β_1 permits agonist-stimulated hydrolysis of PIP_2 in a GTP analogue-dependent manner. The activation of the phospholipase is paralleled by activation of the G proteins, as measured by their binding of the

GTP analogue. These findings demonstrate the complete reconstruction of a signaling pathway for a calcium-mediated agonist from purified components.

The second major research area arose from the observation that many hormones, neurotransmitters, and growth factors break down another cell membrane phospholipid, phosphatidylcholine (PC), in a wide variety of cells. This breakdown yields DAG and phosphatidic acid (PA) and involves two other phospholipases (C and D), which are being purified and characterized. Further work has shown that PC breakdown is regulated by mechanisms involving G proteins, calcium ions, protein kinase C, and tyrosine kinases. The latter two mechanisms are currently being explored in several cell types.

In a related area, the regulation of different isozymes of protein kinase C by certain molecular species of DAG and PA is being studied. This involves the purification of a protein kinase C isozyme that is stimulated by PA but not DAG. These studies should elucidate the functions and mechanisms of this novel signaling system.

Tumor-Suppressor Genes

Andrew P. Feinberg, M.D., M.P.H.—Associate Investigator
Dr. Feinberg is also Associate Professor of Internal Medicine and Human Genetics at the University of Michigan Medical School. He received his B.A., M.D., and M.P.H. degrees from the Johns Hopkins University. He received clinical training at the University of Pennsylvania and Johns Hopkins and did postdoctoral research at the University of California, San Diego, and Johns Hopkins. Before moving to the University of Michigan, Dr. Feinberg was Assistant Professor of Oncology and Medicine at Johns Hopkins.

ONE of the most important areas of cancer genetics is the identification and characterization of tumor-suppressor genes, whose inactivation contributes to cancer. Since almost all genes are present in two copies in the cell, cancer would develop from deletion or inactivation of both copies of these suppressor genes. Inactivation of one copy could be transmitted in families from parent to child. Thus individuals inheriting one nonfunctional copy of the gene would be at increased cancer risk.

Wilms' tumor (WT), a childhood kidney cancer, was one of the earliest models of suppressor gene action. Strong and Knudson showed 20 years ago that WT apparently resulted from two mutations, based on two peaks in the age of onset. Furthermore, it was discovered in the 1970s that some children with WT lack a large portion of chromosome 11 (in band 11p13), which is visible microscopically. Before joining HHMI, I observed (with Eric Fearon and Bert Vogelstein) that more subtle gene deletions can be detected indirectly on chromosome 11 in WTs, through use of restriction fragment length polymorphisms (RFLPs), which can distinguish the maternal and paternal copies of a given gene.

In collaboration with David Schlessinger (Washington University, St. Louis) and Bryan Williams (Hospital for Sick Children, Toronto), our laboratory has now cloned the 11p13 WT gene region in yeast artificial chromosomes (YACs). Interestingly, we found that this region contains multiple genes turned on specifically in developing kidney. At least two of these genes showed reduced or absent expression in approximately half of sporadically occurring WTs, notably those of the same histologic type that occur in children with 11p13 deletions. One of these genes codes for a DNA-binding protein that is mutated in some WTs.

These mutations, however, occur infrequently, and the laboratory is investigating whether reduced expression of the gene may be a more common mechanism for tumorigenesis and whether other genes from this complex may play a role. We are also "retrofitting" WT YACs with a gene that allows growth in mammalian cells, in order to demonstrate directly a tumor-suppressor phenotype of the 11p13 WT gene(s) and to determine by deletion experiments the functional role of the multiple genes from this region.

In addition to the known gene on chromosome 11 that predisposes to WT, we have discovered a second predisposing gene at a different location on the chromosome (band 11p15). This gene appears to be involved in bladder, breast, and lung cancer, as well as WT. Thus WT causation is more complex than investigators had previously believed, and this second WT gene may turn out to be important in common cancers.

Consistent with this gene being a tumor suppressor, the laboratory has applied genetic linkage analysis to map to the same region of 11p15 a cancer-predisposing disorder, Beckwith-Wiedemann syndrome (BWS). Using YACs, the laboratory has now isolated several DNA breakpoints from BWS patients with germline chromosomal rearrangements. This should enable the laboratory to determine whether the BWS gene is indeed the second WT-suppressor gene on the 11p15 band.

Adding to the complexity of WT is the fact that some non-BWS families with hereditary predisposition to WT do not show linkage of this trait to chromosome 11. Recently the laboratory found, in collaboration with former sabbatical member Anthony Reeve (University of Otago, New Zealand), involvement of chromosome 16.

The laboratory has recently developed a novel approach to isolating tumor-suppressor genes directly. Human chromosomes are fragmented into 2–5 million base pair "superfragments" in a way that allows their transfer into any mammalian cell. The advantage of this technique is that it enables one to transfer the giant DNA pieces into a recipient cell and screen directly for a functional gene. It may thus have immediate application to cloning other chromosome 11 tumor-related genes, or more general application to cloning a variety of genes on other chromosomes—genes for which one can currently screen but not select, such as those for cellular aging.

For example, one should be able to exploit the

cancer-suppressing property of tumor-suppressor genes to identify the key chromosomal region containing the gene. In a test of this method, 8 of nearly 100 chromosome 11 superfragments that the laboratory isolated, when introduced into WT cells, suppressed their neoplastic growth. Interestingly, the tumor-suppressing hybrids contain a small portion of 11p15, demonstrating the existence of a suppressor gene on this band. The laboratory is now testing whether these superfragments also suppress important common cancers that involve 11p, such as lung cancer. This novel technique may allow investigators to bridge a gap in cloning methods between chromosome-size pieces (averaging 100 million nucleotides) and YACs (averaging 300,000 nucleotides), and it may have general application to cloning a wide variety of genes.

Another long-standing interest of the laboratory is the role of epigenetic changes (not involving DNA sequence, and thus potentially reversible) in cancer. The laboratory recently identified several CpG islands (DNA sequences rich in cytosine-guanine dinucleotides, often found near actively expressed genes) within the 11p13 WT gene region. Some of these sequences were methylated, a reversible modification of the nucleotide cytosine. Previously CpG islands were only known to be methylated on the inactive X chromosome. The surprising finding of methylated autosomal CpG islands suggests that epigenetic changes may play a role in Wilms' tumorigenesis. We are now testing this hypothesis by examining the islands directly.

We also observed that balanced BWS chromosomal translocations on 11p15 always involve the maternal chromosome, while unbalanced BWS duplications always involve the paternal chromosome, suggesting an epigenetic difference (or imprint) between the two alleles. If such epigenetic changes are found to contribute to inactivation of 11p suppressor genes, this will represent an exciting convergence of our two major interests.

Genetics, Structure, and Function of Histocompatibility Antigens

Kirsten Fischer Lindahl, Ph.D.—Investigator

Dr. Fischer Lindahl is also Professor of Microbiology and Biochemistry at the University of Texas Southwestern Medical Center at Dallas. She began the study of histocompatibility with Morten Simonsen in Copenhagen, Denmark, and received her Ph.D. degree in immunobiology from the University of Wisconsin–Madison. She was a postdoctoral fellow with Darcy Wilson at the University of Pennsylvania, Philadelphia, and with Klaus Rajewsky at the Institute for Genetics in Cologne, West Germany. Before accepting her current position, she was a member of the Basel Institute of Immunology in Switzerland.

HISTOCOMPATIBILITY (H) antigens are cell surface molecules that, when foreign, lead to the rejection of grafted tissues and organs by the vertebrate immune system. Because they form a major obstacle to clinical transplantation, H antigens have been studied for over 50 years. They are complexes of a small peptide ligand and an MHC molecule (encoded by genes of the major histocompatibility complex). A given individual has MHC molecules of a few different kinds, each of which can present to the immune system a large variety of peptides on the surface of cells. These peptides might be derived from proteins produced by intracellular parasites, bacteria, or viruses or by the body's own cells, such as tumor-specific antigens or minor H antigens. The amino acid side chains that line the peptide-binding groove of an MHC molecule determine which peptides that molecule can bind and therefore what antigens can be presented to induce an immune response in the individual with this MHC.

The immune system is capable of recognizing a difference in either of the H antigens' two parts. A difference in the MHC molecule itself will alter many complexes and induce a strong immune response, hence the term "major" H antigen. By contrast, a difference in a peptide alters only one of many kinds of complexes and induces a weaker response, hence the term "minor" H antigen. Unlike the major H antigens, human minor H antigens remain ill defined. In the mouse, however, more than 50 genes that encode minor H antigens have been mapped. Almost every chromosome, including the mitochondrial genome, carries at least one.

The Maternally Transmitted Antigen

In the mouse mitochondrial protein ND1, the sixth amino acid is polymorphic. When cells with one form are transplanted to a mouse with another form, the amino-terminal peptide of ND1 will act as a transplantation antigen, called Mta. This peptide is presented on the cell surface by an MHC class I molecule called M3. M3 binds the ND1 peptide only when the methionine at the end carries a formyl group, and M3 can also bind other peptides with a formyl-methionine. This characteristic of the amino terminus of mitochondrial and bacterial proteins distinguishes them from proteins made in the cytoplasm of mammalian cells. We are collaborating with Michael Bevan (HHMI, University of Washington, Seattle), who is studying the immune response of mice to the bacterium *Listeria*. His group reports that the M3 molecules present a *Listeria* antigen on the cell surface to cytotoxic T cells, which kill the infected cells.

It is of great interest to understand how the formyl-methionine peptides are bound by M3 and which amino acids in M3 are important for this specificity. We cloned the gene for *M3* from mice and rats. The rat and mouse *M3* genes are more similar to each other than they are to the other MHC class I genes of their own species. This is particularly striking in the rat, where the class I genes are otherwise very similar. These results suggest that the specialized function of *M3* evolved long ago in a species from which both rats and mice are descended, and that it has been conserved in both species during their separate evolution, presumably because it is useful in the immune response against bacterial peptides.

We can express *M3* in mouse fibroblasts, where Mta can be detected by killer T lymphocytes. This system allows us to change single amino acids in the protein by introducing mutations at specific sites in the gene. We are currently looking at amino acids 34 and 171, which differ in M3 from the consensus of MHC class I molecules and are near the site where the formyl-methionine of the peptide is expected to lodge. In sequencing naturally occurring variants of the *M3* gene from wild mice, we are learning more about which amino acids are essential for the ability to present the ND1 peptide. Amino acid 95 points straight up in the peptide-binding site, probably right under the variable sixth residue of the peptide. The single change of 95 from leucine to glutamine completely abolishes T cell recognition of Mta. We are collaborating with the laboratory of Johann Deisenhofer (HHMI, University of Texas South-

western Medical Center at Dallas) to produce M3 in amounts sufficient for a structural analysis.

RMA-S Mutant Cells

The heavy chains of MHC class I molecules are not stable in the properly folded conformation at body temperature unless they bind a peptide as well as the β_2-microglobulin light chain. RMA-S mutant cells make MHC class I heavy chains and β_2-microglobulin, but do not display them on their surface. The defect can be circumvented by adding suitable synthetic peptides to the mutant cells, which will then display those MHC molecules that bind the added peptide. Not only classical MHC class I molecules and M3, which were already known to present peptides, are reduced or missing from these cells, but also other MHC class I molecules, such as Qa-1. These molecules are recognized by antibodies or killer T lymphocytes, but their physiological function is not yet clear. The RMA-S cell line shows that they all bind peptides.

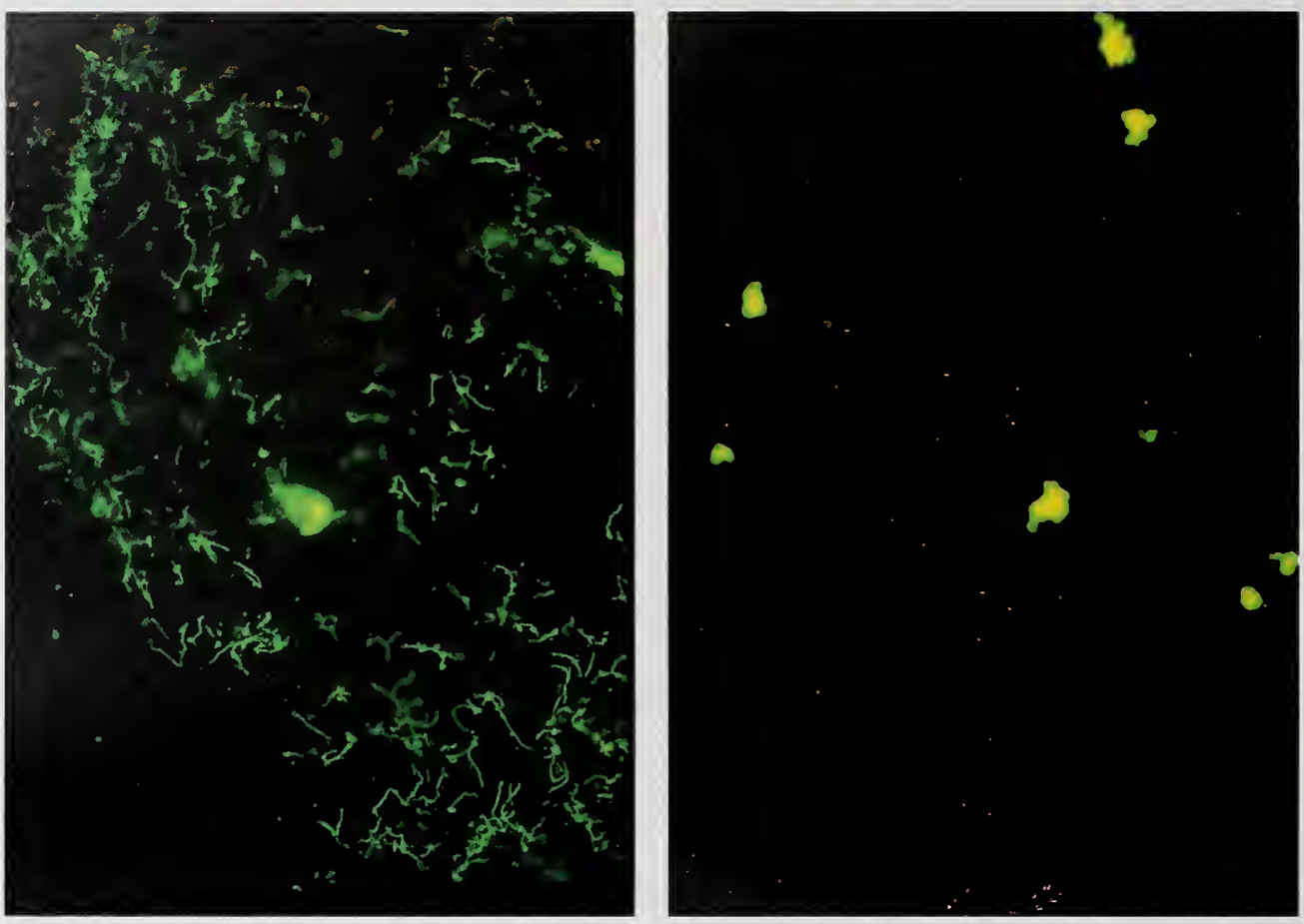
The RMA-S mutation affects one of two MHC genes whose protein products together make a channel to transport peptides from the cytoplasm into the endoplasmic reticulum, where MHC class I molecules fold. Surface display of all such molecules, including Mta and Qa-1, is restored when a normal version of this gene is expressed in the RMA-S cell line. This tells us that the mitochondrial peptide is transported from the cytoplasm into the endoplasmic reticulum by the same mechanism that other peptide antigens use. We have found that the drug oligomycin, which stimulates protein degradation in mitochondria, increases the surface display of both Mta (as expected) and Qa-1 on RMA-S cells, suggesting that Qa-1 can also bind a mitochondrial peptide. We believe that the increased supply of these peptides floods the ineffective transport system and overcomes the defect.

β_2-Microglobulin Polymorphism

Mice are the only mammals in which allelic forms of β_2-microglobulin have been described. Because of its intimate interaction with the peptide-binding parts of MHC class I heavy chains, an allelic difference of β_2-microglobulin that alters amino acid 85 from aspartic acid to alanine can subtly affect the structure of MHC class I molecules and their ability to bind particular peptides. We have found four new alleles of β_2-microglobulin among wild mice of the species *Mus musculus*. These differ by only one or two amino acids relative to the β_2-microglobulin of inbred mice. Three of these wild alleles have valine in position 85 and may therefore alter peptide binding.

We also looked at the β_2-microglobulin gene from the species *Mus spretus*, which has been separated from *Mus musculus* by about 1 million years of evolution. To our surprise, this β_2-microglobulin differed from that of inbred mice by 12 amino acids and had only a single, silent mutation, which affected the DNA sequence, but not the protein. For comparison, rat and mouse β_2-microglobulins differ by 14 amino acids. We would have expected at least an equal number of silent mutations and amino acid changes for the *spretus* form.

Such drastic divergence contrasts with the modest changes seen in other members of the immunoglobulin superfamily and suggests that the molecule has been selected for its differences. The diversification may be part of the speciation process, but it may be limited to molecules that can tolerate change. All changes in the β_2-microglobulin molecule are in loops or on the face away from the MHC class I heavy chain, sparing the face that interacts with the conserved domain of the heavy chains.



Borrelia burgdorferi, *the Lyme disease agent in the gut of deer ticks. Ticks that feed on immunized mice show elimination of the spirochetes (right), whereas ticks that feed on control mice contain numerous spirochetes (left).*

From Fikrig, E., Telford, S., Barthold, S.W., Kantor, F.S., Spielman, A., and Flavell, R.A. 1992. Proc Natl Acad Sci USA *89:5418–5421.*

Genetic Approaches to Immune Function and Tolerance

Richard A. Flavell, Ph.D.—Investigator

Dr. Flavell is also Professor of Immunobiology at Yale University School of Medicine. He received his B.Sc. and Ph.D. degrees in biochemistry from the University of Hull, England, and performed postdoctoral work in Amsterdam and Zurich. Before accepting his current position, Dr. Flavell was first Assistant Professor at the University of Amsterdam, then Head of the Laboratory of Gene Structure and Expression at the National Institute for Medical Research, Mill Hill, London, and subsequently President and Chief Scientific Officer of Biogen Research Corporation, Cambridge, Massachusetts. Dr. Flavell is a Fellow of the Royal Society and a member of other distinguished societies.

MY laboratory has concerned itself for many years with the regulation and function in the immune system of the genes of the murine MHC (major histocompatibility complex). In the mouse these genes are encoded on chromosome 17, and prior work has shown that there are a large number of linked class I–related genes and a handful of class II genes. Class I genes encode a protein of approximately 45,000 molecular weight that is found in association with a small subunit, β_2-microglobulin. Together this complex forms a symmetrical molecule consisting of four extracellular globular domains anchored through the cell membrane with the transmembrane segment and having a short stretch of amino acids that extend into the cytoplasm. Class II molecules achieve a similar symmetry, but with two polypeptide chains, α and β, each of which has two extracellular domains and a transmembrane and cytoplasmic segment.

Both class I and class II gene products serve as recognition elements, which bind antigenic protein fragments and present them to T cells. In the case of class I genes, the presentation is to T cells carrying the CD8 co-receptor molecule. These cells are usually cytotoxic T cells, whose role is to destroy cells that are virally infected. In the case of class II molecules, it is commonly intracellular pathogens that are presented, this time to helper or inflammatory T cells that carry the CD4 co-receptor. Both types of T cells secrete hormone-like molecules called lymphokines, which in turn act on other cell types—for example, on B cells, which are stimulated to multiply and to make antibody.

Class I and class II genes are both regulated *in vivo* by various lymphokines. Of these, the interferons are important and are currently utilized in the therapy of several human diseases. For example, interferon-γ secreted by activated T cells stimulates the synthesis of MHC class I and II molecules and, as a result, presumably renders a cell better able to present antigen and thus to potentiate an immune response. We have taken a genetic approach to attempt to understand how interferon-γ activates the synthesis of these MHC molecules. A new strategy was used to isolate a series of mutant cell lines that are not capable of responding to interferon-γ. These mutants appear to have a series of different defects. In some of these cell lines, the mutations result in a complete loss of the ability of all genes to respond to interferon-γ and even, surprisingly, interferon-α and -β, which were previously believed to use distinct mechanisms. Other cell lines are defective only in the response of the class II and related genes. This genetic approach should help us understand how these important molecules regulate gene expression by elucidating the molecular steps that the cell utilizes to activate genes through interferons.

One important issue in the functioning of the immune system is how the body discriminates its own tissues (self) from foreign components such as pathogens. The process that protects the individual against the destruction of self tissues is known as immune tolerance. Tolerance is generally believed to be established during the production of new T cells in the thymus by a process called negative selection, which is mediated by clonal deletion; that is, self-reactive T cells are destroyed at the site of synthesis. The failure of self-tolerance leads to the autoimmunity that characterizes human diseases such as autoimmune diabetes (insulin-dependent diabetes mellitus [IDDM]) and rheumatoid arthritis.

In the past few years we have been interested in determining the mechanisms of tolerance to those components of the body that are never found in the thymus and therefore pose a problem for tolerance mechanisms operating in the thymus. Transgenic mice can be used to study this process, since the expression of a given gene —and hence the protein encoded by that gene— can be directed to the tissue of choice by linking the gene for the desired protein to the regulatory signals that function in that specific tissue. We have previously performed such experiments by directing the synthesis of MHC class II proteins to the pancreatic β-cells of transgenic mice. In these

experiments the mice were indeed found to be tolerant to the MHC antigens. T cells that would normally react with this MHC product are not eliminated, which is what would be found if the class II antigen is expressed in thymic tissues. Instead, these T cells are present but have been inactivated in some way, such that they are no longer able to respond to the MHC antigen, either in the animal or in test-tube experiments. As a result, no destruction of the pancreatic tissue occurs.

To determine mechanisms of peripheral tolerance to protein antigens, rather than MHC, we made transgenic mice that express the T cell receptor (TCR) reactive with a fragment of the T antigen of SV40 (simian virus 40). We have obtained transgenic mice that express SV40 T antigen in various peripheral tissues in the body, including the β-cells of the pancreas and the secretory exocrine cells of the pancreas that produce digestive enzymes. By crossing the TCR transgenics with mice expressing the antigen, we can determine the immune status to SV40 T antigen. Tolerance is found to the SV40 T antigen provided that expression has occurred prior to the release of T cells from the thymus in the newborn animal. This tolerance results in partial elimination of antigen-specific T cells and clonal inactivation of the remainder. If, however, expression of antigen is delayed, the T cells are not tolerant but instead become activated, with resultant autoimmune destruction of the pancreas.

Delayed expression of antigen therefore seems to place an individual at risk for autoimmunity. However, even in individuals genetically predisposed to become autoimmune, nongenetic factors play an important role. One model suggests that a precipitating event for autoimmunity can be the onset of a local inflammatory response in the tissue, for example, as a consequence of infection. To test this we made transgenic mice that express the inflammatory cytokine TNF (tumor necrosis factor) on the pancreatic islets, which simulates local inflammation. This cytokine expression causes a massive infiltration of T and B lymphocytes, which mimics the infiltrate seen in mouse autoimmune diabetes and causes partial destruction of the insulin-producing β-cells of the pancreas. We are currently determining whether the inflammatory cells present in the TNF transgenic mice are specific for pancreatic antigens, as is the case in true IDDM.

We also study the response of the immune system to the spirochete *Borrelia burgdorferi*, which causes the notorious inflammatory disease, Lyme disease. We showed last year that mice vaccinated against an outer surface protein of *Borrelia* (OspA) are protected against infection by *Borrelia burgdorferi*. We have now shown that protection can also be mediated by a second protein (OspB) but not by the flagellar antigen, even though a potent antibody response is obtained to this protein.

For such a vaccine to work it must be effective against most, if not all, strains of the infectious agent and be effective against a natural challenge. Since the natural vector is the deer tick, we also infected laboratory mice with *B. burgdorferi* by placing infected ticks on these mice. Encouragingly, mice vaccinated with either OspA or OspB were protected; furthermore, spirochetes in the gut of ticks feeding upon vaccinated mice were eliminated, whereas they were unaffected in nonvaccinated mice. This shows an additional potential benefit of the vaccine, which is to eliminate spirochetes from the vector. We are now considering strategies for the eradication of the spirochetes in the wild population of ticks by a related approach.

Biophysical Genetics of Protein Structure and Folding

Robert O. Fox, Ph.D.—Associate Investigator

Dr. Fox is also Associate Professor of Molecular Biophysics and Biochemistry at the Yale University School of Medicine. He received his B.S. degree in biochemistry from the University of Pittsburgh and his M.Phil. and Ph.D. degrees in molecular biophysics and biochemistry from Yale University while working with Frederic Richards in the area of x-ray crystallography. He carried out postdoctoral studies at Yale University in protein engineering with Nigel Grindley and studied protein folding using NMR spectroscopy with Christopher Dobson at Oxford University as a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. Before moving to Yale, Dr. Fox was Assistant Professor in the Department of Cell Biology at Stanford University Medical School.

ALTHOUGH the information that directs the folding of a protein molecule into a defined three-dimensional structure is genetically encoded, the mechanisms and pathways of the folding process are poorly understood. One approach to this problem is an analysis of partially structured folding intermediates, combined with a mutational analysis. We use nuclear magnetic resonance (NMR) spectroscopy and chemical methods to probe for structural and kinetic intermediates in the folding process.

Many polypeptide sequences adopt a common folded motif, but they frequently differ in the detailed arrangement or conformation of structural elements in ways that are functionally significant. Certain loops of the immunoglobulins (antibodies) are examples. We are using staphylococcal nuclease as a model protein system to understand how the amino acid sequence of a secondary structural element dictates its detailed conformation in the context of a folded protein molecule. We combine a number of methodologies in these studies, including x-ray crystallography, NMR spectroscopy, and molecular biology.

Mapping Structure in the Unfolded State of Proteins

Protein molecules in the unfolded and molten globule states are often more compact than would be expected for a true random-coil conformation. If this conformational bias is toward that of the folded structure, it may explain the rapid rate at which proteins fold. We have developed a chemical approach to map close contacts between a variable-reporter residue site and all other residues of a protein chain in these states. This approach is being used to investigate staphylococcal nuclease variants, nuclease fragments, and the molten globule state of myoglobin. A polar chelator has been designed and synthesized that can be specifically attached to a cysteine residue engineered into the protein chain. When this chelator is loaded with iron and the reaction is initiated with a reducing agent, hydroxyl radicals and other reactive oxygen species are generated; these in turn cleave peptide bonds at positions in the protein chain in proximity to the chelator. The cleavage sites can be determined by peptide mapping and protein sequencing. The reagent cleaves native proteins at a number of solvent-accessible sites close to the site of attachment. The reagent has also been used to map the proximity of several residues of $\gamma\delta$ resolvase to its DNA-binding sites.

Analysis of Protein Folding Using NMR Spectroscopy

Protein molecules are generally thought to adopt a final tertiary structure where all backbone and side chain conformations and tertiary contacts are within local energy minima. Several well-refined protein molecules display exceptions to this view, where significant strain is induced at a particular point in the structure, usually involving a residue in the enzyme active site. Examples include *cis*-peptide bonds, eclipsed side chain rotamers, and energetically unfavorable van der Waals contacts. How a protein imposes the stress needed to favor locally strained conformations remains unclear.

Staphylococcal nuclease provides an excellent system to examine the relationship between stress and strain in a globular protein. Residues 115–118 adopt a type VIa β-turn, containing a *cis*-peptide bond, forming one wall of the nucleotide-binding pocket of the active site. Although the conformation of the Lys116-Pro117 peptide bond of this β-turn is predominantly in the *cis*-configuration, the equilibrium between *cis*- and *trans*-conformers can be monitored by NMR spectroscopy. Site-directed mutants within this type VIa β-turn have been examined by NMR spectroscopy and x-ray crystallography. Limiting the backbone conformational space available to the residue preceding the *cis*-proline contributes to the stress favoring the strained *cis*-peptide bond conformation. We have begun a collaboration with Axel Brünger (HHMI, Yale University)

to simulate this system using molecular dynamics techniques.

A structural analysis of early protein-folding intermediates has been difficult because of their transient nature. Amide hydrogen exchange has been used by others to examine the appearance of hydrogen-bonded protein-folding intermediates for a number of proteins. We have begun to use this technique to examine early folding intermediates of a staphylococcal nuclease variant. Partial protection of amides occurs early in the folding process, despite the long overall folding time for nuclease.

Genetic Analysis of a β-Turn

A sharp change in the trajectory of a polypeptide chain between secondary structure elements in a globular protein has been defined as a reverse turn or β-turn. These structures occur in a number of defined geometric types and frequently contribute side chains to the active site of the enzymes, such as staphylococcal nuclease, or the combining site of binding proteins, such as the immunoglobulins. We wish to determine the sequence requirements for the formation of different β-turn types to better understand the detailed structure of globular proteins and to define design principles for protein engineering.

We have developed a genetic approach to determine which amino acid sequences are consistent with a particular β-turn structure in staphylococcal nuclease. Each member of our gene library contains a unique sequence at this β-turn. Only a small fraction of the sequences examined are consistent with an enzymatically active and stable protein in *Escherichia coli*. The β-turn under consideration is well removed from the active site, suggesting that the modulation in the observed enzyme activity is due to changes in the stability of the protein. There are strong biases in the amino acids occurring at each position in the β-turn. Recently, a statistical analysis of these data has led to a predictive model for this β-turn type in all globular proteins. This approach may be useful in defining other sequence–secondary structure relationships. Results of this and related experiments should provide insight into the relationship between amino acid sequence and structure required for the rational design and engineering of protein molecules.

Structural Studies of Trypanosome Calmodulin

Calmodulin serves as a calcium-dependent regulatory subunit of a variety of cytoskeletal proteins and cytoplasmic enzymes, including a number of protein kinases. The novel structure of rat calmodulin was determined in the laboratory of Charles Bugg. The protein is composed of two largely helical domains, each with two calcium-binding sites, separated by an extended solvent-exposed α-helix. We have obtained crystals of calmodulin from *Trypanosoma brucei rhodesiense*, in collaboration with Curtis Patton (Yale University). The crystal structure in progress should serve as the basis for rational drug design against this and related organisms.

Molecular Basis of Genetic Diseases and Chromosome Mapping

Uta Francke, M.D.—Investigator

Dr. Francke is also Professor of Genetics and Pediatrics at Stanford University School of Medicine. She received her M.D. degree from the University of Munich, Germany, trained in pediatrics at Los Angeles Children's Hospital, and carried out postdoctoral research and clinical training in medical genetics at the University of California (Los Angeles and San Diego). Before moving to Stanford, Dr. Francke was Professor of Human Genetics and Pediatrics at Yale University School of Medicine.

THE genetic maps of humans and the mouse are undergoing rapid growth and development. Through the efforts of many laboratories, including ours, several hundred homologous genes have been mapped in both species, and over 60 chromosome regions have been delineated that contain conserved groups of genes. Thus it has become possible, after mapping a gene in one species, to predict the location of its homologue in the other. Comparative mapping information is used to evaluate the possibility of a mouse mutation being a true model of a human genetic disorder.

Our laboratory is employing *in situ* hybridization, with multicolor nonradioactive detection of chromosomal signals, as well as somatic cell genetic approaches, to locate cloned genes of known function on human and mouse chromosomes. We are using this information to define candidate genes for human inherited disorders or for phenotypic mutations in mice and to further delineate regions that contain homologous genes in both species. Our goal is to identify genes involved in producing phenotypic abnormalities in chromosomal imbalance syndromes and in inherited disorders, to understand their function, and to devise precise diagnostic tests and rational treatment strategies.

Search for Genes Involved in Diseases

In collaboration with the laboratory of Eric Shooter, Stanford University, we have mapped a peripheral myelin protein gene, identified by others as having a growth arrest function, to mouse chromosome 11 and human chromosome 17. It thus became a candidate gene for involvement in the mouse mutation *Trembler* and the inherited human neuropathy Charcot-Marie-Tooth disease, type 1A. Point mutations in this gene were indeed demonstrated in *Trembler* mice, and the work on the human disorder is in progress.

In collaboration with Stuart Leff and Tim Donlon, also at Stanford, we have assigned a gene whose product is involved in mRNA-processing (splicing) events to the region of human chromosome 15 that is commonly deleted in patients with the Prader-Willi syndrome. In this deletion syndrome, hypotonia, hypogonadism, mental retardation, and obesity due to lack of appetite control are associated with often submicroscopic (micro-) deletions of region 15q11.2-q13. Since no gene of known function has yet been mapped to the smallest deletion overlap region, our assignment of a gene to this region makes it a candidate for contributing to the microdeletion phenotype. It is well established that in Prader-Willi syndrome the chromosome with the deletion is paternally derived, while the maternally derived homologous genes do not appear to be expressed (imprinting). If we can show that our candidate gene is also imprinted on the maternally derived chromosome, it would support the hypothesis that this gene contributes to the deletion phenotype.

While the classical forms of X-linked–recessive progressive muscular dystrophy (Duchenne and Becker types) are due to deletions or mutations in the dystrophin gene on Xp21, there is a distinct autosomal recessive form of muscular dystrophy, affecting both sexes and clinically resembling Duchenne muscular dystrophy, for which the defect is unknown. We have collected and studied several families with more than one affected individual. The dystrophin-like gene on chromosome 6 has been excluded as a candidate, and linkage to other chromosomal sites is being tested. We are also collaborating with Kevin Campbell (HHMI, University of Iowa, Iowa City), whose laboratory has characterized and isolated a dystrophin-associated complex of glycoproteins located at the sarcolemma that interacts with dystrophin. We have begun to map genes for these proteins to chromosomal sites as a prerequisite for testing them as possible candidates for involvement in autosomal childhood-onset progressive muscular dystrophy.

Search for Mutations in Candidate Genes

In order to find mutations in a candidate gene, we are employing screening methods in which the gene is amplified in small portions of DNA of

an obligate heterozygote. The amplified DNA duplexes are screened for the presence of base differences by denaturing gradient gel electrophoresis (DGGE) or single-strand conformational polymorphism (SSCP) studies. Amplified exons that appear to have base differences, identified by altered melting behavior or conformational properties, are sequenced to reveal the precise base differences.

In a rare form of inherited dwarfism, the Laron syndrome, individuals are unresponsive to growth hormone, lack the activity of a specific growth hormone–binding protein in serum, and have low levels of insulin-like growth factor I. The growth hormone receptor is a likely candidate gene for mutations leading to this disorder. The receptor gene consists of 10 exons that are amplified individually from flanking primers. In a study of 38 Laron syndrome individuals from a highly inbred population in the mountains of southern Ecuador, we identified a base substitution in the extracellular domain of the receptor that does not alter the encoded amino acid but creates a new donor splice site. The resulting mutant mRNA is deleted for 24 bases, which predicts a protein lacking eight amino acids in a highly conserved region of the molecule. The location of the deleted eight amino acids does not involve the hormone-binding or receptor dimerization sites. The most likely effect of this mutation is a poorly folded, unstable protein that might degrade rapidly. Further studies are under way to test this prediction.

The same strategy is now being pursued to find mutations in the growth hormone receptor gene in other patients with Laron syndrome who are from different ethnic and racial populations. The Ecuadorean growth hormone receptor mutation represents the first instance of a disease-causing base substitution that acts by creating a new donor splice site and leads to its exclusive use while the original site is unaltered.

The gonadotropin-releasing hormone (GnRH) is produced in specialized cells in the hypothalamus, travels to the pituitary gland, and causes the release of the gonadotropic follicle-stimulating hormone (FSH) and luteinizing hormone (LH). We previously mapped the gene to the short arm of human chromosome 8. In the mouse, partial deletion of this gene causes a recessive phenotype, termed *hypogonadal,* leading to lack of sexual development. In human families with inherited incomplete sexual development due to low levels of gonadotropin hormones, the GnRH gene is a candidate for involvement, as in the mouse model. We have studied a family with three affected siblings that show a sequence polymorphism in the signal peptide of the GnRH gene, but have found that this polymorphism does not cosegregate with the disease phenotype, thus excluding a GnRH gene defect in this particular family.

The Marfan syndrome is an autosomal dominant condition that involves abnormalities in elastic tissue of the aorta and the fibrils that suspend the optic lens and in connective tissue of many organs, including bones, tendons, and lung. The estimated incidence is 1 in 10,000, and clinical severity is highly variable. Dissection and rupture of the aorta is a life-threatening complication. During the past year, linkage studies in other laboratories have pinpointed the site of the Marfan gene on human chromosome 15, and coincidentally the gene encoding fibrillin, the major microfibrillar protein, was also mapped to the same location. Subsequently, a mutation in the fibrillin gene was documented in two unrelated Marfan patients.

Stanford University is a center for Marfan diagnosis, treatment, and research. Over the last three years, we have established a broad-based research program that involves a detailed clinical and genealogical database and the collection of blood and tissue samples from affected persons and family members. We hypothesize that a mutation in the fibrillin gene causes the Marfan phenotype by producing an abnormal protein that when incorporated into the microfibrils causes their instability.

We are now screening the mRNA produced by skin or aortic fibroblasts from patients with classical Marfan syndrome and related connective tissue disorders. After reverse transcription of the 10,000-bp fibrillin mRNA, small overlapping sections are amplified and screened for point mutations by DGGE and SSCP studies. Our goal is not only to develop a molecular diagnostic test for a disorder that is clinically often difficult to diagnose, but also to gain insight into the relationships between particular types of mutations and the phenotypes they produce.

Since there are fibrillin-related genes elsewhere in the genome and many connective tissue disorders that overlap the Marfan phenotype to some degree, this will be a large fruitful field of investigation.

Molecular Biology of Obesity and Diabetes

Jeffrey M. Friedman, M.D., Ph.D.—Assistant Investigator

Dr. Friedman is also Associate Professor and Head of Laboratory at the Rockefeller University. He received his B.S. and M.D. degrees from the Renssalaer Polytechnic Institute–Albany Medical College. After completing a residency in internal medicine at Albany Medical College and a gastroenterology fellowship at Cornell University Medical College, he enrolled in the graduate program at Rockefeller, where he received his Ph.D. degree in molecular biology.

EXTENSIVE studies of humans and other organisms have suggested that body weight, body composition (percent body fat), and food intake are under strict physiological control. The "set point" hypothesis holds that both the intake and expenditure of energy are physiologically regulated in the individual to maintain a predetermined body weight. Implicit in this hypothesis is the notion that signal molecules reflecting the nutritional state are synthesized in the periphery and sensed by brain centers whose appropriate response is set to stabilize body weight.

Studies in which the rodent brain has been selectively lesioned suggest that this feeding control center resides, at least in part, in the hypothalamus. However, the site of synthesis of the molecules that signal nutritional state is unknown, though fat cells or cells of the gastrointestinal tract have been proposed. Knowledge of the site of synthesis and the molecular nature of signals affecting the control of appetite and body composition could have important implications for our understanding of nutritional disorders.

To learn more about these signaling mechanisms, we have been taking a variety of approaches, both genetic and molecular, to study the role of specific gene products in the control systems.

Molecular Basis of Obesity in *ob/ob* and *db/db* Mice*

If one wishes to understand the basis for differences in the complicated system of energy homeostasis, there are a number of experimental advantages to studying mutant mice, including the ability to control for environment and to set up genetic crosses. For these reasons, we have begun a study of mice carrying recessive mutations that result in profound obesity. At least four obesity-causing mutations are available: *obese* (*ob*), *diabetes* (*db*), *fat* (*fat*), and *tubby* (*tub*). In each case, a mouse becomes obese because of a single-gene defect. We have focused on the *ob* and *db* mutations for several reasons. The mutant mice become very obese, often three times normal weight; and the obese phenotype in mutant animals, as in humans, appears to result from both increased food intake and diminished energy expenditure. Furthermore, Douglas Coleman at the Jackson Laboratory has suggested that *ob* mice lack a circulating factor that suppresses appetite and that *db* mice, which are unable to respond to this factor, may lack its receptor.

Current techniques in molecular genetics, such as Southern blots and chromosome walking, make it possible to clone genes such as *ob* and *db* whose function is known on the basis of a mutant phenotype but whose gene product is unknown. This approach, called positional cloning, utilizes restriction fragment length polymorphisms (RFLPs)—genetic markers defined by specific cloned pieces of DNA. These markers define genetic differences in inbred mouse strains. The first step in attempts to clone a mutant gene utilizes genetic crosses between normal and obese (or diabetic) mice. Inheritance of the obese phenotype is compared in individual animals with that of individual RFLPs, which if inherited along with the obese phenotype, are said to be linked genetically to the obesity gene.

By performing this analysis on several thousand mice with several dozen different RFLPs, we have been able to identify a series of DNA probes that are very tightly linked to the *ob* and *db* genes. RFLPs linked genetically to these mutations are in physical proximity to *ob* and *db* and can be used as starting points to characterize the adjacent DNA and clone the mutant genes.

In the case of *ob* mice, we first used three different RFLPs to generate a detailed genetic map around the *ob* locus. The genes represented were the *met* oncogene carboxypeptidase A, which codes for a pancreatic enzyme; the *irp* gene; and the cystic fibrosis gene. Similarly, the mouse *db* gene has been mapped relative to RFLPs for interferon-α and a complement gene. We have also found that the rat obesity mutation *fatty* (*fa*) is also flanked by the genes for interferon-α and

* Studies aimed at cloning the mutant obesity genes *ob* and *db* from mice have been supported by the National Institute of Diabetes and Digestive and Kidney Diseases.

complement. These data suggest that mutations in the same gene can cause obesity in two different rodent species and raise the possibility that similar mutations can cause obesity in humans. We are endeavoring to test this possibility, in collaboration with Rudolph Leibel, by using the interferon-α and complement genes to characterize human families with a high incidence of obesity.

To identify other RFLPs that are more tightly linked than the probes mentioned above, we have used the technique of chromosomal microdissection, in which small slices of individual chromosomes are dissected and cloned. Separate libraries have been made from proximal chromosome 6, where *ob* maps, and from mid chromosome 4, where *db* maps. Two probes have been isolated from the chromosome 4 library that flank *db* and are about 1 cM apart. Similarly, two probes that flank *ob* and are about 0.2 cM apart were isolated from the chromosome 6 library. A genetic distance of 0.2 cM corresponds to about 400,000 base pairs of DNA.

We are currently attempting to clone the DNA near these probes by using techniques such as pulsed-field gel electrophoresis and cloning of large fragments in yeast artificial chromosomes (YACs). YACs of 300,000 to 1 million base pairs have already been isolated for each of the probes that flank *ob*, and the search for the gene in these artificial chromosomes should begin shortly. The cloning of these genes should further our understanding of the mechanisms that control food intake and body weight.

Polygenic Inheritance of Type II Diabetes

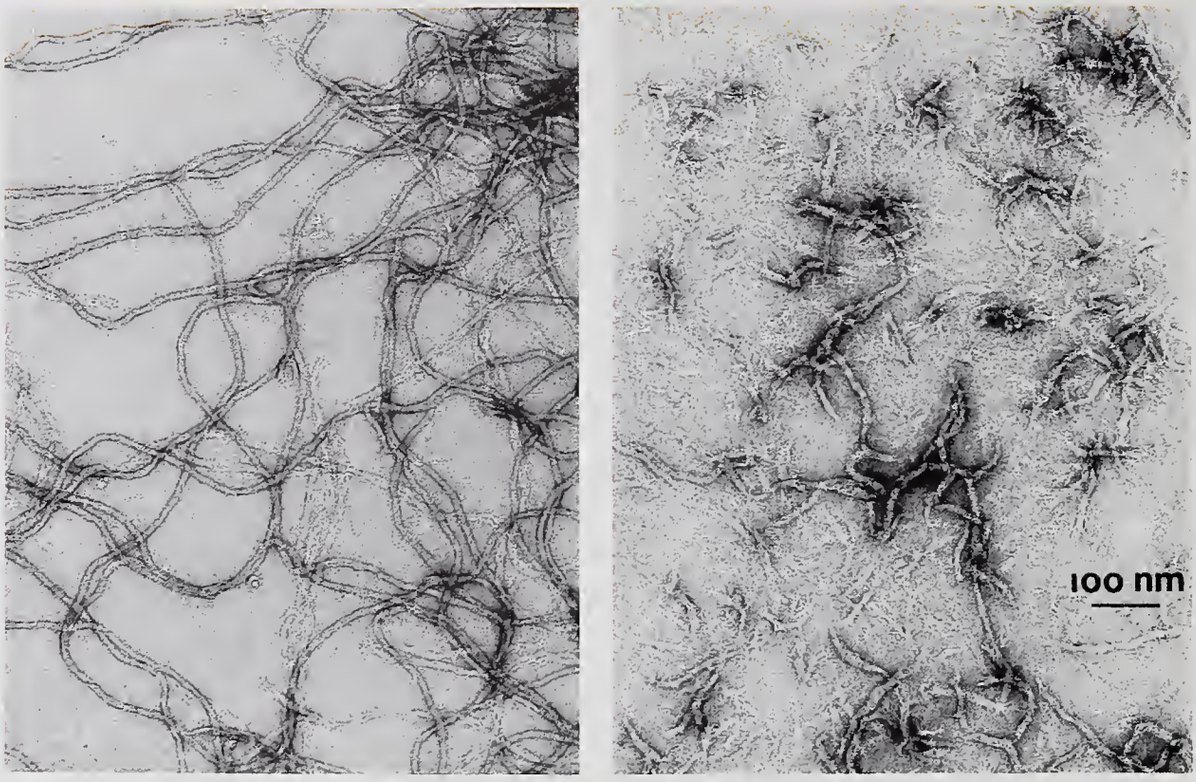
The mice carrying the *ob* and *db* obesity mutations develop diabetes of adult onset that is quite similar to type II diabetes in humans. However, differences in the type II phenotype are seen in mutant mice, depending on the strain carrying the mutation. C57BL/6J *ob/ob* mice develop a mild insulin-resistant diabetes with high levels of plasma concentration of glucose and insulin. In contrast, DBA/2J mice develop a severe diabetes characterized by very high plasma levels of glucose and relatively low plasma insulin concentration. These data suggest that genetic differences in insulin sensitivity and output from pancreatic β-cells can influence the severity of the diabetes in genetically obese mice of different inbred strains.

Further analyses of these data have suggested that the genetic differences are a result of polygenic inheritance—genetic variation in several genes. Advances in molecular genetics now make it possible to dissect polygenic traits (such as diabetes in mice) into single gene components. This requires that a collection of genetic markers, such as RFLPs, be available for genotyping the diabetic animals. To facilitate the analysis of polygenic traits, we have helped Eric Lander and William Dietrich (Massachusetts Institute of Technology) generate a linkage map of the mouse genome, using a new type of genetic marker known as the SSR (simple sequence repeat). These markers greatly facilitate genetic mapping experiments. We are now analyzing obese progeny of various genetic crosses with these SSR markers to identify novel genes that predispose *ob* animals to diabetes.

Regulation, Function, and Expression of Cholecystokinin in Human Tumors

The hormone cholecystokinin (CCK) was originally found in the small intestine by virtue of its ability to stimulate gallbladder contraction and pancreatic secretion in response to feeding. High levels of CCK have also been found in neurons of the mammalian brain, where it functions as a neurotransmitter. The first demonstration that CCK could affect behavior was reported by Gerry Smith and Dick Gibbs, who showed that peripherally administered CCK had an appetite-suppressing effect on rats. It has also been demonstrated that CCK antagonists increase feeding behavior in rodents. These observations suggest that the regulation and function of this gene is important in the control of appetite.

At present, we are using a variety of techniques to explore CCK's function(s). One of our first objectives was to determine whether any human tumors associated with weight loss overexpress this peptide, and we therefore began screening tumor cell lines for CCK production. Several pediatric tumors were found to synthesize the appetite depressant, including peripheral neuroepithelioma (a rare nerve tumor that usually develops in the chest wall), Ewing's sarcoma of bone, and rhabdomyosarcoma (a malignant muscle tumor). These data suggest that measurement of the CCK levels in the blood may be of diagnostic and prognostic value in the management of these tumors. Studies are being initiated to ascertain whether overproduction of CCK in patients with these tumors is associated with excessive weight loss.



Test-tube filaments made from genetically engineered human epidermal keratins. On the left are filaments from normal keratins; on the right, filaments from keratins with a single amino acid substitution (arginine to cysteine). This point mutation was discovered in a patient with epidermolysis bullosa simplex, a disease involving abnormalities in the keratin filament network.

From Coulombe, P.A., Hutton, M.E., Letai, A., Hebert, A., Paller, A.S., and Fuchs, E. 1991. Cell *66:1301–1311. Copyright © 1991 by Cell Press.*

Regulation of Keratin Expression During Differentiation and Development in Human Skin

Elaine Fuchs, Ph.D.—Investigator

Dr. Fuchs is also Professor in the Departments of Molecular Genetics and Cell Biology and of Biochemistry and Molecular Biology at the University of Chicago. She received her B.S. degree in chemistry from the University of Illinois and her Ph.D. degree in biochemistry from Princeton University, where she studied with Charles Gilvarg. Her postdoctoral research was done with Howard Green at the Massachusetts Institute of Technology. Dr. Fuchs counts among her honors the R.R. Bensely Award from the American Association of Anatomists.

THE long-range objective of our research is to understand the biochemical mechanisms that operate and regulate the expression of human genes during development and differentiation in skin. Present knowledge of the biochemistry of human skin and its diseases is limited. Although dermatologists have always directed their interest toward human skin biology and skin diseases, the field of molecular biology has only recently approached a level of understanding that permits the complex biochemistry of human skin to be explored. A major factor facilitating such studies is the ability to grow human skin cells in tissue culture, including epidermal cells, dermal fibroblasts, melanocytes, and dermal papillae cells. The recent development of technology for targeting foreign genes to transgenic animals' skin has provided a valuable *in vivo* model system for the study of genetic skin diseases and skin cancers.

Much of our research on human skin has focused on the epidermis, which comprises about 20 cell layers whose outermost is the skin surface. Only the inner or basal layer of the epidermis is truly living and undergoes DNA synthesis and cell division. Under an influence as yet unidentified, a basal cell ceases to divide and makes a commitment to differentiate terminally. As the cell moves outward to the skin surface, it undergoes a variety of morphological and biochemical changes. The most pronounced of these is the production of a dense network of keratin filaments, which are tough, resilient protein fibers. Many skin diseases of the epidermis, including psoriasis and basal and squamous cell carcinomas, involve a malfunctioning of the differentiative process that is frequently associated with some abnormality in the production or organization of these filaments. Our investigation is focused on the regulation of the expression of keratin proteins and their genes in human epidermis and in epidermal cells differentiating in tissue culture.

The keratins are a group of 10–20 related proteins (40–70 kDa) that form the 10-nm keratin filaments in the cytoplasm of epidermal cells. Only a subset (typically 2–6) of keratins are ever expressed at one time. As a normal epidermal cell differentiates, it changes the subset of keratins it makes. In addition, the cell increases its keratin synthesis, leaving the fully differentiated epidermal cell with 85 percent of its total protein as keratins. In diseases of the skin involving epidermal hyperproliferation, including psoriasis and squamous cell carcinomas, a new subset of keratins not normally made in the epidermis is produced, and can be diagnostic.

A coordinated genetic and biochemical approach is necessary to determine the regulation of the multiple keratins and to decipher their structural and functional roles in the differentiating epidermal cell. A number of years ago, we showed that expression of different subsets of epidermal keratins is due to changes in the synthesis of different mRNAs. We used DNA recombinant technology to show that these mRNAs are encoded by about 20 different genes of two distinct types. Type I encodes small keratins (40–53 kDa); type II, larger keratins (53–67 kDa).

Keratins are expressed as specific pairs of type I and II proteins. The basic subunit of keratin filaments is a heterodimer, composed of the two types. Approximately 20,000 heterodimers form a single 10-nm filament; the assembly process is energy independent and does not appear to require auxiliary proteins or factors. Using DNA sequencing, we determined the amino acid sequences for several keratin pairs. The cytoskeletal architecture of keratin filaments may be specifically tailored to suit the particular structural needs of each epidermal cell at various stages of differentiation and development.

To determine the details of the filament assembly process and to investigate the interactions of keratin filaments with other proteins and organelles, we used deletion and site-directed mutagenesis to alter the coding sequences of K5 and K14, the pair expressed in the living cells of the epidermis. We generated substantial quantities of keratins for filament assembly studies, using genetic engineering to overexpress wild type and mutant human keratins in bacteria, which do not have keratin. We purified these keratins and isolated and examined the consequences of mutations and deletions on keratin filament assembly *in vitro*.

In the past few years, we identified those sequences involved in filament elongation and those that are more important for lateral associations. In addition, we used gene transfection of human tissue culture cells to examine the dynamics of filament assembly *in vivo* and to determine how abnormal expression of keratins in malignant and hyperproliferating epidermal cells might change their biology and cytoskeletal architecture. The results of these studies have begun to yield valuable insights into the complex assembly process of keratin filaments and the function they perform in providing a tough skin surface.

We have also been interested in determining whether natural mutations in human epidermal keratin genes might lead to genetic skin diseases. We had previously shown that elevated expression of epidermal keratin genes is a relatively late event in development and that certain keratin mutants have a deleterious effect in cultured cells, disrupting the endogenous keratin filament network. In the past two years, we have made transgenic mice and used epidermal keratin promoters to target expression of some of these dominant mutant keratins to the epidermis.

Unexpectedly, the transgenic mice exhibit a phenotype resembling that of humans with epidermolysis bullosa simplex (EBS), a class of blistering skin diseases that are dominant, sometimes life-threatening, and of previously unknown etiology. Mice expressing K14 mutants that severely disrupt filament assembly exhibit severe blistering over body trunk regions—lesions resembling those of Dowling Meara EBS, the severest form; while mice expressing K14 mutants that mildly perturb the network exhibit blistering only over paws, as in the mildest form of EBS, or Weber-Cockayne, where patients show blistering predominantly on their hands and feet. Thus we were able to demonstrate that multiple mutations in a single gene, namely K14, can give rise to most if not all forms of EBS, thus strongly suggesting that these diseases may be genetically linked. Further studies enabled us to demonstrate that skin blistering in EBS is due to the fragility of basal epidermal cells caused by an abnormal keratin filament network, thereby compromising the mechanical strength of this cell layer.

The similarities between the phenotypes in mice and humans prompted us to focus on this disease. In the past year we characterized the K14 and K5 mRNAs, genes and proteins from two patients with Dowling Meara EBS. Both patients have point mutations in an amino acid of the K14 protein that we had previously shown to be critical for filament assembly. When we engineered and tested the two mutations in our wild-type K14 gene, we verified that these defects are responsible for the phenotype of the EBS patients. It will take much work to develop better diagnostic and therapeutic tools for EBS, but elucidating the genetic basis is an important first step.

While EBS is a disease of defects in the coding portions of keratin genes, there are other skin diseases that appear to arise from abnormalities in the control of epidermal genes or genes that influence growth and/or differentiation. As a prerequisite to investigating the bases for these types of diseases, we are 1) analyzing the molecular mechanisms underlying the differential expression of epidermal keratin genes and 2) utilizing our K14 promoter to make transgenic mice that overexpress various regulatory factors—e.g., transforming growth factor-α (TGFα), epidermal growth factor (EGF), cytokines, hormone receptors, and proto-oncogenes in the epidermis.

In the past few years, we have identified proximal and distal domains that act synergistically to regulate expression of the human K14 gene. Our goal is to identify the sequences and transcription factors involved. Analysis of the factors controlling the cell's major structural genes should lead us to the factors determining keratinocyte fate. Unraveling the nature of promoter and enhancer sequences involved in regulating epidermal genes will be important not only for understanding epidermal development but also for targeting products to the epidermis. In the last year, we used the K14 promoter/enhancer to engineer transgenic mice that overexpress TGFα, the major autocrine growth factor of the epidermis. Our results have shown that TGFα plays an important role in controlling the thickness of the epidermis and that certain features of TGFα overexpression are similar to those of psoriasis, a hyperproliferative skin disease.

The K14 promoter/enhancer, so effective in transgenic technology, should also be useful for drug therapy. Because epidermal cells can be removed from a patient, cultured *in vitro*, and grafted back, it should be possible to introduce foreign genes, driven by the K14 promoter/enhancer, into the cultured epidermal cells prior to grafting. Such techniques have potentially powerful applications for future medical research.

The Molecular Basis of Viral Replication and Pathogenesis

Donald E. Ganem, M.D.—Associate Investigator

Dr. Ganem is also Professor of Microbiology and Immunology and of Medicine at the University of California, San Francisco. He received an A.B. degree in biochemistry from Harvard College and the M.D. degree from Harvard Medical School. Following his clinical training in infectious diseases, he did postdoctoral research training in the laboratory of Harold Varmus at UCSF.

OUR laboratory studies the molecular mechanisms by which pathogenic human viruses infect the host and cause disease. We are especially interested in those viruses that produce persistent infections and engender chronic pathology. Several fundamental questions underlie our work: How are persistent infections established and maintained? What factors regulate viral replication and spread? How does the persistent presence of virus evoke disease?

Most of our work centers around the human hepatitis B virus (HBV) and its animal homologues. HBV is a small DNA virus that replicates principally in liver cells (hepatocytes) and produces acute and chronic type B hepatitis. Most initial infections are transient: following a relatively brief period of liver injury (hepatitis), the immune system eliminates virus from the liver and invokes lasting immunity. However, 5–10 percent of infections are not successfully eliminated. In these cases, viral replication persists in the liver for the life of the host, evoking various degrees of chronic liver injury that can lead to premature death from liver failure. Most strikingly, persistence of viral replication for several decades enormously increases the risk of liver cancer.

There are many reasons to be interested in this remarkable infection. First is its great public health significance. Worldwide, there are over 250 million chronic HBV carriers. In large regions of Asia and Africa, 10–15 percent of all human inhabitants are persistently infected by the virus. Second, the virus replication cycle is distinctly unusual: replication of the viral DNA genome is accomplished via reverse transcription of an RNA intermediate. The HBV life cycle is thus a permuted version of the retroviral life cycle, and a fuller understanding of its details allows instructive comparisons to be made with cognate steps in retroviral replication. More importantly, unraveling the mechanism of HBV replication should identify new potential targets for antiviral therapy. Finally, the nature of the link between chronic HBV replication and hepatocellular carcinoma represents one of the great unsolved problems in human cancer biology, and a deeper understanding of HBV persistence is expected to provide new clues to its solution.

Our studies of HBV replication have examined many different steps in the viral life cycle. The initial step in all viral infections is binding of the virus to the cell surface receptor and its entry into the cell. Little is known about this step in HBV infection, but we have begun to explore it in a convenient animal model, the duck hepatitis B virus (DHBV). By using recombinant DHBV surface proteins to look for cellular proteins that will bind them, we have identified a single host glycoprotein that will interact with DHBV particles with high affinity and specificity. Such molecules are good candidates for the receptor, and we are actively attempting to clone the gene for this host protein to allow its more-detailed characterization.

Next, internalized virus particles must be delivered to the nucleus, where their DNA genome is transcribed. This is perhaps the most poorly understood of all reactions in virology. For no virus has the cellular machinery involved in this essential step been identified. Recently, we discovered that drugs affecting the integrity of cellular microtubules block this step, implying that these structures are involved in the intracellular transport of subviral particles. This raises the exciting possibility that such particles might also be used to identify host proteins that recognize them and perhaps bind not only virus particles but normal cellular components (e.g., organelles) to cytoskeletal motors for transport within the cell.

Once in the nucleus, viral genes are transcribed into RNA to be transferred to the cytoplasm for translation. Following translation, progeny particles assemble in the cytoplasm. In this remarkable reaction, viral RNA is packaged into subviral particles along with the viral reverse transcriptase. Surprisingly, this RNA packaging process requires the participation of the reverse transcriptase itself. To understand this key step better, we are currently attempting to reproduce this RNA recognition *in vitro*, using recombinant reverse transcriptase and cloned viral RNA. Sim-

ilar *in vitro* reconstruction reactions are being attempted for the reverse transcription reaction itself. This enormously complex reaction generates the duplex DNA genome from its RNA precursor. By fully characterizing the molecules that catalyze the reaction and the intermediates through which it proceeds, we hope to identify novel targets for drugs that would be useful in interrupting virus growth.

The development of liver cancer during persistent infection is another major focus of interest. For this we study two relatives of HBV: the woodchuck hepatitis virus (WHV), which is potently oncogenic in its normal host, and the ground squirrel hepatitis virus (GSHV), a less-efficient carcinogenic stimulus in woodchucks. Our studies thus far suggest fundamental differences in the oncogenic pathways employed by these closely related viruses. WHV-induced tumors frequently (30–45 percent) contain integrated WHV genomes adjacent to the N-*myc* oncogene, whose expression is thereby activated. Such insertional activation is not apparent in GSHV-induced hepatomas. We are currently looking for other genes that are regularly disrupted by WHV integration, with the view that such loci may define other proteins centrally involved in the control of normal hepatocyte growth.

Second Messengers and Cell Regulation

David L. Garbers, Ph.D.—Investigator

Dr. Garbers is also Professor of Pharmacology at the University of Texas Southwestern Medical Center at Dallas. He received his B.S. degree in agriculture and his Ph.D. degree in biochemistry at the University of Wisconsin. His postdoctoral research was done at Vanderbilt University. Before assuming his present position, Dr. Garbers was Professor of Pharmacology and of Molecular Physiology and Biophysics at Vanderbilt University School of Medicine.

THE focus of the research in this laboratory centers on the mechanisms by which cells communicate with each other—specifically the mechanisms by which sea urchin or mammalian spermatozoa detect signals from the egg. These studies have turned out to be applicable to somatic cells and have led to the identification of a new cell surface receptor family in humans and other mammals. Members of this family serve as receptors for molecules that regulate blood pressure, as well as a large number of other physiological processes.

Around 1981 our laboratory reported the purification of a peptide that can stimulate sperm motility. This peptide was derived from media in which sea urchin eggs had been allowed to stand. Subsequently it was demonstrated that different species of sea urchins contain different peptides and that the molecules from one species do not necessarily stimulate the sperm cells from another. In later research it became clear that sperm cells detect higher concentrations of peptide and swim toward them. Since the highest concentrations are around the egg, the sperm cell swims directly toward the egg under normal conditions.

This laboratory set out to determine how the cell detects the egg peptide. We found that a particular protein on the sperm plasma membrane specifically bound the egg peptide. This protein appeared to serve as the detector or receptor molecule, but how it signaled to spermatozoa that a specific egg peptide had been bound was not clear.

To help resolve the question, we purified the receptor protein on the plasma membrane. It was identified as the enzyme guanylyl cyclase, which catalyzes the formation of cyclic GMP, a small molecule that causes a change in the behavior of many different cells. It seemed possible that the receptor is situated with part of it outside the cell, where it could bind the egg peptide, and the other part inside, where it forms cGMP. The cGMP thus formed would then serve as a signal that egg peptide is being detected, lending greater speed and direction to the spermatozoon.

To provide evidence that the membrane receptor protein is in fact guanylyl cyclase, we isolated complementary DNA clones for this enzyme. Such clones allow one not only to predict the primary structure of a protein, the receptor in this case, but to direct protein synthesis in quantity. Unfortunately, the sea urchin sperm receptor was not formed in the proper manner, and research continues on its expression.

Under appropriate conditions, DNA will bind (hybridize) to closely related DNA. Therefore sea urchin DNA was used to determine whether mammals contain a related protein that might serve as a receptor. Clones containing complementary DNA were isolated from rat brain. The DNA sequence revealed that a rat brain guanylyl cyclase is a component of the plasma membrane, with approximately one-half of the protein outside and one-half inside the cell. The intracellular region of the sea urchin sperm enzyme is very similar to the part of the rat brain enzyme inside the cell. In the regions outside the cell, however, the two proteins show little similarity. This would be expected if guanylyl cyclase serves as a cell surface receptor for peptides and the peptide of mammals is different from that of the sea urchin. That is, the detector part of the molecule would need to change to recognize a different peptide, but intracellular regions of the receptor could remain unaltered.

Subsequent binding studies demonstrated that the cloned rat brain guanylyl cyclase could specifically bind and be activated by certain peptides synthesized in the heart and brain. These atrial natriuretic peptides (ANPs) regulate blood pressure, as well as various other physiological events.

Since sea urchin sperm cells respond to peptides that do not resemble ANP, other animals may be expected to contain yet other peptides that interact with guanylyl cyclase. In addition, multiple membrane forms of guanylyl cyclase may exist within mammals. We have used the DNA that encodes the ANP receptor to determine whether other guanylyl cyclase receptors exist. Another receptor with properties similar to the one described above has been identified. This

guanylyl cyclase receptor is similar to the first ANP receptor within intracellular regions but is only 43 percent identical in the extracellular, ligand-binding region. It now appears that a peptide other than ANP regulates this receptor.

We subsequently discovered the existence of a third guanylyl cyclase receptor. This receptor binds small peptides released from various bacteria that cause acute diarrhea. This form of diarrhea (often referred to as traveler's diarrhea) is prevalent in infant humans and young domestic animals. The receptor has the same general features as the two receptors described above but has a markedly different amino acid sequence within the toxin-binding region. Whether a regulatory molecule not of bacterial origin normally exists that binds to this receptor is not yet known.

During the past year, we have obtained evidence to suggest that at least two other guanylyl cyclase receptors exist. The hormones that normally regulate these receptors have not yet been identified. It seems, therefore, that different cells contain unique guanylyl cyclase receptors, which allow them to respond to specific hormonal signals with an increased production of the messenger molecule cGMP. The cells then change their behavior. For example, one receptor may control smooth muscle relaxation and be principally involved in the regulation of blood pressure, another may be involved in neural functions, and yet another may regulate epithelial cell secretion. These results also suggest that drugs specific to different forms of the guanylyl cyclase receptors may prove clinically relevant.

Molecular Genetics of the Major Histocompatibility Complex

Jan Geliebter, Ph.D.—Assistant Investigator

Dr. Geliebter is also Assistant Professor and Head of the Mammalian Molecular Genetics Laboratory at the Rockefeller University. He received his Ph.D. degree in microbiology and immunology from the State University of New York, Downstate Medical Center. He was a postdoctoral fellow and research associate in the laboratory of Stanley Nathenson at the Albert Einstein College of Medicine, Bronx, New York.

THE immune system functions to rid the body of foreign objects such as bacteria, viruses, tumors, and transplants. The portion of such matter that is recognized as foreign by the immune system is called an antigen. Antigens that are found on cells are "presented" to the immune system by cell surface molecules called histocompatibility molecules (also called HLA molecules in humans and H-2 molecules in the mouse). Histocompatibility molecules are able to bind antigenic fragments of, for example, viruses, and stimulate white blood cells (lymphocytes) to attack the virus-infected cell, thereby limiting the spread of infection. Without these antigen-presenting molecules the host would be unable to mount an immune response against pathogens and would not survive.

Different H-2 molecules can bind and present different types of antigens. Because inbred mice have about three different types of H-2 molecules on their cells, they can bind and present a large, but limited, number of antigens to the immune system.

To ensure the survival of the species, it is beneficial that many varieties of H-2 molecules be present in the population. In this way there will always be some portion of the population that will mount an immune response to a given antigen. Indeed, an extraordinary number of different histocompatibility molecules have been found in almost all species investigated. In humans, the large variety of HLA molecules ensures our survival but is the major obstacle confounding tissue transplantation. Our interest lies in the genetic mechanism that generates the different histocompatibility genes in mice and other species.

The *H-2* genes of the mouse are part of the larger major histocompatibility complex class I multigene family. This gene family also contains genes that are structurally similar to *H-2* genes and have unknown functions. The genetic mechanism that generates variety in *H-2* genes is the microrecombination process, which reassorts DNA among *H-2* genes and other related class I genes. By substituting small segments of class I gene sequences into *H-2* genes, the microrecombination process can create new H-2 molecules that have different antigen-presenting capabilities, thereby expanding the immune responsiveness of the population.

Our interest is to understand better the mechanism underlying the microrecombination process. This process has previously been studied by identifying microrecombinant mice that differed from their otherwise identical siblings by altered *H-2* genes. Since microrecombinant H-2 molecules elicit skin graft rejection, these studies were accomplished by skin graft testing thousands of mice. The rejection of a skin graft by a sibling mouse signaled an alteration in H-2 molecules. These labor-intensive studies found that, on the average, one microrecombinant mouse was detected for every 5,000 skin grafts performed.

To gain further insight into the microrecombination process, we are using *in vitro*–engineered constructs to detect microrecombinant *H-2* genes. We have constructed a fusion gene in which β-galactosidase sequences replace two cytoplasmic exons for the K^b gene. The fusion protein can be detected by staining for β-galactosidase activity, which is manifested as blue-colored cells. We have also site-directed two in-frame termination codons in the K^b gene at positions that undergo frequent microrecombinations. This prevents the expression of β-galactosidase. β-Galactosidase expression can be rescued by a microrecombination with a linked class I gene, *Q4*, that recombines away the termination codons. Thus microrecombinations can be scored as blue cells.

This microrecombination construct, once introduced into a variety of cell types, will help determine the microrecombination frequencies of different cells. Its introduction into transgenic mice will help determine frequencies in germ cells. Data from previous studies indicate that microrecombinations occur in female germ cells. We also hope to determine if they occur in sperm cells as well and at what frequency. Some strains of mice may undergo microrecombinations more

frequently than others. Placing this construct onto different genetic backgrounds may help to determine microrecombination frequencies in different mouse strains and perhaps identify critical parameters in the process. These studies will contribute to our understanding of the genetic processes that control the evolution and ultimately the function of the mammalian immune system.

Although sequence diversity and polymorphism are the hallmark of genes, region genes are characterized by sequence conservation among alleles and limited polymorphism. The lack of polymorphism among these genes has been suggested to preclude an immunological function for the gene products. We have identified a region gene whose sequence differs greatly between alleles of the C57BL/6 and C3H mice. The sequence differences between the two alleles are manifested in both scattered and clustered nucleotide substitutions. The clustered substitutions are similar to those observed in the microrecombinations that diversify genes and may reflect past microrecombination events with *H-2* and other *Qa* region genes. These data may provide the first evidence that *Qa* genes can be recipients in the microrecombination process.

Polymerase chain reaction (PCR) analysis has indicated that this *Qa* gene is transcribed in some strains of mice. We are currently engaged in an in-depth analysis of the transcription, translation, and cell surface expression of this gene and its product to ascertain its function. This gene is polymorphic in at least three strains of mice, and the analysis of other strains is under way. The diversity and polymorphism of this gene suggest an immunological function for its product, perhaps the first *Qa* gene to be ascribed such a function.

The Decoding Code in mRNA

Raymond F. Gesteland, Ph.D.—Investigator

Dr. Gesteland is also Professor of Human Genetics at the University of Utah School of Medicine and Professor of Biology and Adjunct Professor of Bioengineering at the University of Utah. He received his B.S. degree in chemistry and his M.S. degree in biochemistry from the University of Wisconsin. He earned his Ph.D. degree in biochemistry from Harvard University, where he studied with J. D. Watson. Dr. Gesteland was an NIH postdoctoral fellow at the Institute de Biologie Moléculaire, Geneva, Switzerland. He served as Assistant Director and Investigator at Cold Spring Harbor Laboratory, New York, before assuming his present responsibilities.

THE genetic code precisely specifies the identification of each of the 20 amino acids from the set of 61 triplets in mRNA, with three triplets dedicated to termination. As a ribosome moves progressively along the mRNA deciphering it, the contiguous triplets contain all the information necessary to specify amino acid sequence. However, there is another code, also encrypted by the linear order of bases in mRNA, that carries information about the mechanism of translation.

This code specifies how the common genetic code should be implemented for each mRNA. In some cases the additional information may be quite simple, consisting of instructions that tell the ribosome where to start and where to stop decoding. In other cases this secondary code radically changes the decoding process so that individual mRNAs use a different version of the genetic code, altered either in the meaning of certain codons or in the linear readout mechanism. This set of instructions could be called the "decoding code," in that it operates on the genetic code. Different decoding codes individualize mRNAs.

A decoding code operating during the deciphering of a specific mRNA often results in a violation of the conventional genetic code. This can occur with a set frequency, so that the same mRNA produces two different protein products; but the frequency can also depend on metabolic state, thus providing another point at which gene expression is controlled.

For example, the ribosomal machinery can be instructed to read a nontriplet number of bases at one site so that an alternate reading frame will be used from that point on. Or a proportion of the ribosomes can be instructed to jump from one site to another on the same mRNA so that noncontiguous codons will be read out. Or the meaning of specific codons can be altered so that a ribosome reads a stop codon as an amino acid, or even as a new amino acid that is not a member of the conventional set of 20. Each of these unusual events is specified by the decoding code information.

The diversity of schemes for encrypting information for the decoding code in individual mRNAs is just beginning to be appreciated. In some cases, rather simple sequences are involved; in others, sequence dictates complex RNA structures.

The signals in mRNA that carry the instructions for programmed ribosomal frameshifting usually include a shift site where the frame changes and a stimulator sequence that greatly increases the shift site's efficiency. The shift site consists of 4–7 nucleotides that allow a ribosomal-bound tRNA decoding in the first frame to slip forward or backward by one nucleotide to get into the new frame. The stimulator information can be upstream of the shift site, as in the case of the *RF2* gene in *Escherichia coli*, where 5–6 nucleotides need to pair directly with a specific sequence in ribosomal RNA in order to stimulate the shift. Or it can be downstream, as in the case of many retroviruses and retroviral-like genes where the specific interactions are less clear and the mRNA contains secondary and tertiary structures that somehow stimulate frameshifting at the upstream shift site. The combination of shift site and stimulator sequences constitutes the decoding code that tells how the standard genetic code should be corrupted for translation of a specific message.

There is only one compelling case of ribosomal jumping, and here the decoding code is complex. The mRNA for gene 60 of bacteriophage T4 has a gap of 50 nucleotides that is very efficiently bypassed by ribosomes. The mRNA sequence elements involved include both structural information in the gap and upstream information in the form of the amino acid sequence of the growing peptide chain that must interact with the same part of the complex.

Other versions of the decoding code reprogram stop codons to specify amino acids. Some retroviruses (e.g., murine leukemia virus, MuLV) make a fusion protein from two in-frame genes separated by a UAG stop codon, which in this case is decoded as glutamine. The decoding code that specifies this includes a downstream element whose

secondary and tertiary structures are crucial. In the handful of known cases of proteins having the unusual amino acid selenocysteine, a UGA stop codon is reprogrammed to be decoded as this 21st amino acid. The decoding code clearly includes structural elements that are downstream, in one case 150 nucleotides away in the nontranslated part of the message.

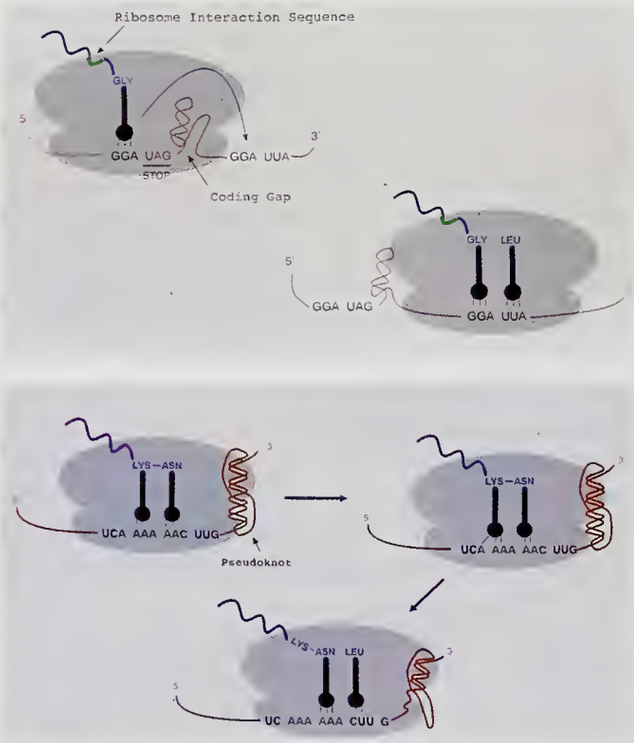
An emerging theme of the decoding code is mRNA secondary and tertiary structural elements. Stem-loop structures can often be identified by inspection of sequence, and evidence for their participation can be obtained by making disruptive and restorative mutations. The functional stems can be short in some contexts, long in others, and can be near the site of action or well downstream.

In an increasing number of cases, the crucial structure is more complicated, involving a pseudoknot where a sequence downstream of a stem-loop folds back to pair with the loop of the stem, resulting in two stems and two loops that are intertwined. Physical-chemical properties of model pseudoknots show that the two stems are coaxial, with considerable stability coming from the extended base stacking that is achieved.

To test the biological importance of each of the bases of these complicated structures requires construction of a very large number of mutants, and analysis of the resulting structural changes is so far minimal. In some cases the actual sequence of the proposed stems is unimportant so long as base-pairing is maintained and the sequence of the loops is not crucial. In fact, these criteria are used to define the importance of a stem or pseudoknot. In at least one case (MuLV stop codon read-through), a few bases in the loop region of a pseudoknot are crucially important.

Structures in ribosomal RNA are beginning to be delineated clearly, largely from evolutionary comparisons of sequence. We are still quite ignorant, however, about the complexity of structures in mRNA. At least for the elements involved in the decoding code, there is a biological assay for function.

We are even less clear about how these structures in mRNA alter the decoding process. The simplest view—that mRNA structures merely cause a ribosomal pause allowing alternative reactions—is almost certainly naive. But whether the structures bind protein factors, interact with ribosomal proteins, pair with ribosomal RNA, or even interact with other parts of mRNA is completely untested. Investigation of these possibilities in the context of the decoding code will be revealing.



Decoding of the genetic code. Top: Model of a ribosome traversing the 50-nucleotide coding gap in mRNA of bacteriophage T4's gene 60, illustrating the involvement of upstream peptide sequences and sequences in the gap.

Bottom: Model of a ribosome reading a stop codon in mRNA of a murine leukemia virus, stimulated by a downstream pseudoknot structure.

Research of Norma Wills, Robert Weiss, Dianne Dunn, and John Atkins in the laboratory of Raymond Gesteland.

Molecular Analysis of Proteins Involved in Human Disease

Mary-Jane H. Gething, Ph.D.—Investigator

Dr. Gething is also Professor of Biochemistry at the University of Texas Southwestern Medical Center at Dallas. She received her bachelor's and Ph.D. degrees from the University of Melbourne. After holding research fellowships in Cambridge, England, with Brian Hartley and in London with Michael Waterfield, she joined the scientific staff of the Imperial Cancer Research Fund, London. After that she was a senior staff investigator at the Cold Spring Harbor Laboratory, New York, before going to the Southwestern Medical Center.

OUR studies on experimental models of human disease involve three systems: 1) human tissue-type plasminogen activator (t-PA), a serine protease involved in fibrinolysis, tissue remodeling, and metastasis; 2) the hemagglutinin of influenza virus, which is being used to develop models of autoimmune disease in transgenic mice; and 3) the tumor-suppressor protein p53 and its interaction with cytosolic stress-70 proteins.

Role of t-PA in Thrombolysis and Metastasis

Many normal and abnormal biological processes that require extracellular proteolysis, including thrombolysis, tissue remodeling, and metastasis, are mediated by plasminogen activators that cleave plasminogen to the active protease plasmin. One such activator, t-PA, is the principal thrombolytic agent in the circulation, and its elevated expression is thought to be linked to increases in the metastatic potential of some types of tumor cells, including malignant melanomas.

The t-PA protein is composed of a number of independent structural domains. The finger domain and an epidermal growth factor (EGF)-like domain are involved in the initial, high-affinity binding of t-PA to fibrin, while stimulation of t-PA activity requires secondary, lower-affinity interactions of fibrin with either of two kringle domains of the molecule. The binding of t-PA to specific clearance receptors on hepatic cells also involves sequences within the finger and/or EGF-like domains. Finally, a specific inhibitor, plasminogen activator inhibitor-1 (PAI-1), interacts with the active site in the carboxyl-terminal catalytic domain.

Although the three-dimensional structure of t-PA has not been solved, we have been able to model all the domains through use of known structures of homologous domains in other proteins. Site-directed mutants using these proposed structures have provided information about the role of individual amino acid sequences of the protein, and variant enzymes have been generated that are efficient, fibrin-stimulated plasminogen activators but are resistant to inhibition by a variety of serpins, including PAI-1, or do not bind to the t-PA receptor(s) involved in clearance of the enzyme in the liver. Because these mutant enzymes should have an extended effective life in the circulation, they may have significant potential for use in thrombolytic therapy of patients with myocardial infarction.

The variant enzymes are also being utilized to test the role of t-PA in metastasis of malignant melanoma cells. Transgenic mice expressing the T antigen oncogene from simian virus 40 (SV40) under the control of the mouse tyrosinase promoter develop ocular melanoma with high frequency. These tumors are transplantable to nontransgenic animals but are not metastatic. We are currently developing additional lines of transgenic mice that express wild-type or inhibitor-resistant forms of t-PA from the same tyrosinase promoter. Crossing of the Ty-Tag and Ty-tPA transgenic mice will reveal whether an increased level of t-PA production in melanoma cells would result in increased metastatic potential.

Transgenic Models of Immune Tolerance and Autoimmune Disease

We are using transgenic mice to study the development of immunological responses to the hemagglutinin (HA) of influenza virus. RIPHA mice, which express HA from the rat insulin II promoter/enhancer only in the β-cells of the pancreas, promise to provide a valuable model for the study of immune tolerance and autoimmune diabetes. These mice display no physiological problems until 4–5 months of age. After that, increases in the blood glucose levels begin to appear, and shortly thereafter these mice develop a severe hyperglycemia that is responsive to insulin.

In the RIPHA-33 line, both male and female animals can develop an immune response to HA and other antigens of pancreatic β-cells. However, male mice become hyperglycemic with much higher frequency (45 percent) than female mice (5 percent). Recent studies have shown that the amount of fat in the diet is an important regu-

lator of diabetes in RIPHA-33 mice and that the sex bias may reflect a decreased ability of male mice to control their blood glucose levels. Current studies focus on a comparison of the RIPHA model with mice of the non-obese diabetic (NOD) line and on the role played by cytokines, such as interleukin-1, tumor necrosis factor, interferon-γ, and interleukin-6, during the progression of the disease.

Folding and Assembly of the Tumor-Suppressor Protein p53

Until recently it was widely assumed that the folding and oligomerization of newly synthesized polypeptides and their subsequent molecular rearrangements are spontaneous processes that do not require the intervention of other cellular proteins. However, it is now apparent that members of the stress-70 protein family are "chaperones" intimately involved in facilitating protein folding and assembly within prokaryotic and eukaryotic cells.

Our previous studies of the interaction of BiP, the endoplasmic reticulum stress-70 protein, with newly synthesized membrane and secretory proteins has led to an understanding of the role of stress-70 proteins in stabilizing unfolded or partly folded polypeptides in a form competent for further folding and oligomeric assembly. Current studies focus on the role of hsc70, the cytosolic stress-70 protein, in modulating the structure or activity of mutant forms of the tumor-suppressor protein p53.

Other workers have shown that p53 plays a role in regulation of normal cell growth and that it binds to a number of viral-transforming proteins, including SV40 T antigen. Mutations in p53 convert the protein to an oncogenic form capable of co-operating with another oncogene, activated *ras*, to transform cells. These mutant forms of p53 appear to be altered conformationally, as indicated by loss of T antigen recognition, altered reactivity with monoclonal antibodies, and co-translational binding to hsc70. We are analyzing the interaction between p53 and hsc70, using the fd bacteriophage expression system. By screening p53-peptide–expressing bacteriophage for binding to hsc70, with subsequent enrichment and amplification in bacteria of those expressing hsc70-binding epitopes, the region(s) in p53 recognized by hsc70 will be identified.

Signal Transduction Pathways in B Lymphocytes

Sankar Ghosh, Ph.D.—Assistant Investigator

Dr. Ghosh is also Assistant Professor of Immunobiology and of Molecular Biophysics and Biochemistry at Yale University School of Medicine. After receiving his Ph.D. degree from the Albert Einstein College of Medicine, Bronx, New York, he conducted postdoctoral work at the Whitehead Institute in Cambridge, Massachusetts, under the supervision of David Baltimore.

THE focus of research in our laboratory is to understand the signals that regulate B cell development and the pathways through which they are transduced. We propose to address these questions by studying the regulation of developmental-stage–specific expression of the immunoglobulin κ light-chain gene. Primarily, we wish to find the signal that initiates κ gene expression during B cell development and determine whether expression of the gene is necessary for rearrangement of the κ locus.

The expression of the κ gene in mature B cells is driven by an enhancer located in the intron between the Jκ and Cκ genes. The activity of this enhancer depends on the binding of a transcription factor, called NF-κB, to a specific site in the element. In keeping with its role as a regulator for developmental-stage–specific expression of the κ gene, NF-κB can only be detected in an active form in the nucleus of mature B cells. However, it is present in the cytoplasm of pre–B cells and other cell types as an inactive precursor, bound to an inhibitory protein called IκB.

NF-κB activity can be induced by treatment of cells with agents such as phorbol esters or bacterial lipopolysaccharide. These agents activate signal transduction pathways that result in the modification of IκB and the subsequent dissociation of the NF-κB:IκB complex. The free NF-κB then enters the nucleus and activates gene expression. Therefore the inducible system of NF-κB behaves like a second messenger, transducing cell surface signals to the nucleus.

To understand the change in IκB that results in the dissociation of the NF-κB:IκB complex, we established an *in vitro* system to look at the involvement of protein kinases in the activation process. The results strongly suggested phosphorylation of IκB as a key event in the activation of NF-κB. However, further analysis of the activation process was hampered by the lack of reagents, such as antibodies, or clones for the genes encoding NF-κB. Therefore we undertook to clone the genes; the information we obtained has resulted in significant new insights into the function of these proteins.

NF-κB is composed of two subunits of 50 kDa (p50) and 65 kDa (p65) and exists as a heterodimer. Cloning the cDNA encoding p50 revealed that it was actually synthesized as a larger protein of 105 kDa, which was processed to the mature 50-kDa size. The full-length protein did not bind to DNA, but removal of the carboxyl-terminal region revealed the DNA-binding activity. However, the most interesting feature of the clone was the homology for over 300 amino acids in the amino-terminal half (i.e., the region encoding p50) to the oncogene v-*rel*, its cellular counterpart c-*rel*, and the *Drosophila* morphogen *dorsal*.

Subsequent cloning of the p65 subunit revealed a similar *rel*-homology region, which contains both DNA-binding and dimerization domains. In fact, with the recent cloning of other *rel*-homologous proteins, a picture has begun to emerge in which the different members associate with one another to form transcription factors that can activate different genes.

The homology between NF-κB and *dorsal* also has interesting implications for understanding the biology of both proteins. The developmental morphogen *dorsal* is responsible for the establishment of the dorsoventral polarity in *Drosophila* embryos. The basis for its action is the establishment of a gradient of dorsal protein that is progressively localized to the nucleus. The cytoplasmic localization of *dorsal* was genetically determined to be due to the action of a gene called *cactus*, which behaves like IκB, and recent cloning of the gene has shown that it is a homologue of mammalian IκB.

The signal for the nuclear localization of *dorsal* goes through a pathway involving eight upstream maternal-effect genes. However, for those of us studying NF-κB, what has been of great interest is the finding that one of these upstream genes, *toll*, is a membrane receptor with a cytoplasmic domain that is homologous to one of the mammalian interleukin-1 (IL-1) receptors. This is significant because, among the different interleukins, only IL-1 can activate NF-κB. Therefore the two genes that lie between *toll* and *dorsal*

(i.e., *pelle* and *tube*) are probably components of the signal transduction pathway that result in the modification of *cactus*. The challenge for the future is to find the analogues of *pelle* and *tube* in mammalian cells, and we are exploring different approaches in our attempt to elucidate the signal transduction pathway that results in the constitutive activation of NF-κB in mature B cells.

The next major advance was the cloning of IκB, and once again the cloning revealed a sequence homology that is important for IκB function. In this case the sequence elements are the ankyrin repeats that are found in a diverse group of proteins, including the p105 precursor of NF-κB p50, *Drosophila* Notch, yeast cdc10, and the human erythrocyte ankyrin protein. Because all of these proteins are either cytosolic or membrane associated, the initial idea was that ankyrin sequences mediate cytosolic retention of proteins by binding to cytosolic or membrane structures. However, a nuclear transcription factor was recently found to contain these sequences, necessitating a change in our idea of the role of ankyrin sequences. A more comprehensive model envisions them as domains responsible for protein-protein interaction. We also found that pp40, a protein associated with v-Rel and c-Rel in chicken lymphoid cells, is a homologue of mammalian IκB. This observation, along with the finding that *cactus* is also homologous to IκB, suggests that the NF-κB/*rel/dorsal* family has evolved over millions of years and used the same general principles for regulating the activity of these positive-acting transcription factors.

One of the most interesting questions about NF-κB is its regulation during B cell development. NF-κB is an inducible cytosolic protein in pre–B cells but changes into a constitutively active nuclear protein in mature B cells. There can be at least two simple explanations for this: either a transcriptional shutoff of IκB synthesis or a constitutive modification of IκB in mature B cells. Preliminary results indicate that transcription of the major form of IκB is not shut off in mature B cells, but the IκB that is synthesized is somehow inactivated. We are trying to determine how IκB activity is regulated in mature B cells. We also want to determine the signal responsible for the developmental transition of a pre–B cell to a mature B cell.

The other major focus of our research is to understand the regulation of κ gene expression in plasma cells. Unlike mature B cells, plasma cells express large amounts of the immunoglobulin chains, including the κ chains, which form complete immunoglobulin molecules that are secreted. The high-level expression of the κ gene is not driven by the intronic enhancer; instead, an enhancer located 9 kb downstream of the κ gene appears to be the dominant element in plasma cells. Therefore κ expression appears to be controlled by a unique dual enhancer system during development: the intronic enhancer drives low-level expression in mature B cells, while the 3′ enhancer is responsible for high-level expression in plasma cells. The questions we wish to address in this system are, What is the protein(s) responsible for the activity of the 3′ enhancer in plasma cells? What is the signal that activates the 3′ enhancer? and finally, What role does crosslinking of the surface immunoglobulins play in this process? We would also like to study the surface immunoglobulin complex on mature B cells that includes proteins such as mb-1 and B-29 and elucidate the signal transduction pathway that leads to the activation of the 3′ enhancer.

Molecular Genetics of Blood Coagulation

David Ginsburg, M.D.—Associate Investigator

Dr. Ginsburg is also Associate Professor in the Departments of Internal Medicine and Human Genetics at the University of Michigan Medical School. He received his B.A. degree in molecular biophysics and biochemistry from Yale University and his M.D. degree from Duke University School of Medicine. His postdoctoral research training was done in the laboratory of Stuart Orkin at the Children's Hospital, Harvard Medical School. While in Boston, Dr. Ginsburg was also Instructor in Medicine at Brigham and Women's Hospital, Harvard Medical School.

THE major research activities of my laboratory concern two important blood clotting proteins, von Willebrand factor and plasminogen activator inhibitor-1, and their associated human diseases. In addition, we are applying molecular tools to the study of bone marrow transplantation.

von Willebrand Factor

One major function of von Willebrand factor (vWF), which is an important part of the body's blood clotting system, is to serve as a bridge between blood platelets and the wall of an injured blood vessel, thereby helping to control bleeding. vWF is also the carrier for factor VIII, the missing substance in patients with hemophilia. Abnormalities in vWF result in von Willebrand disease (vWD), the most common inherited bleeding disorder, occurring in 1–3 percent of the general population. Over 20 different types of vWD have been described.

Our aim is to understand how the various parts of the vWF protein work in the body and how they interact with other factors in the blood clotting system. Recently we have made considerable progress in identifying the defects within the vWF gene that are responsible for vWD. In addition to aiding in the diagnosis and management of vWD, this information has provided important insights into the function of vWF.

Over 90 percent of patients with type IIB vWD have one of four specific defects, all within a small region of the vWF gene critical for its interaction with blood platelets. By introducing one of these defects into the DNA of tissue culture cells, we have shown that this single change is responsible for the type IIB variant. In similar studies of type IIA vWD, we have identified another set of defects clustered in a different region of the vWF gene. We have introduced these defects into cultured cells and have shown that type IIA vWD may be due to abnormalities in the process whereby vWF is manufactured inside the cell. In studies of a patient whose vWF is unable to bind factor VIII, we identified a single change in the gene that has helped to pinpoint the region of vWF responsible for factor VIII transport. We have also identified in some patients with type III vWD (the most severe form) a defect in the ability to copy the vWF gene into normal messenger RNA.

Despite considerable progress, type I vWD, the most common variant, remains a mystery. We have recently begun a new project to study the molecular basis for a disease in mice that closely resembles type I vWD of humans. Surprisingly, it appears that the defect in the mouse may be in a gene other than vWF. If we are successful in identifying the mutant gene in the mouse, our findings should be directly transferable to the study of human type I vWD. Through these studies, we hope to expand our understanding of vWF, to advance our ability to diagnose and classify vWD, and eventually to improve the medical treatment for this common human disorder. The work on von Willebrand factor has been funded in part by a grant from the National Institutes of Health.

Plasminogen Activator Inhibitor-1

The fibrinolytic system is the body's mechanism for breaking down blood clots. This system must be precisely balanced with the clot-forming system, since an imbalance can result in unwanted blood clotting or uncontrolled bleeding. The protein that turns on the fibrinolytic system is plasminogen activator. Its activity is controlled by a regulator protein, plasminogen activator inhibitor-1 (PAI-1).

Synthetic plasminogen activators, such as recombinant tissue-type plasminogen activator (t-PA) and urokinase, are now used in patients to dissolve blood clots, particularly in the early stages of a heart attack when a major blood vessel to the heart muscle has become blocked. There is also increasing evidence that patients with abnormally high blood levels of PAI-1 (disrupting the normal clot-dissolving activity of natural plasminogen activator) are at particularly high risk for heart attacks and other diseases due to increased blood clot formation. Thus an understanding of the structure and function of PAI-1 and its interaction with plasminogen activators is

of great significance in the study of the fibrinolytic system and its role in a number of important diseases.

We have established a system for directing bacteria to make synthetic PAI-1 in the test tube. This recombinant PAI-1 has all of the important properties of the natural protein. We have made over 170 variants of PAI-1 in which a small part of the molecule has been changed. Studying the effects of these small changes, which can vary considerably in their ability to regulate the different types of plasminogen activators, has advanced our understanding of how PAI-1 functions and how it interacts with other parts of the blood clotting system. These observations might eventually aid in the design of new drugs for the treatment of bleeding and clotting diseases. This work has been funded in part by a grant from the National Institutes of Health.

We have recently identified the molecular defect in our first patient with complete deficiency of PAI-1. This nine-year-old girl has a moderately severe bleeding disorder. Studies of her DNA demonstrated complete inactivation of the gene, accounting for the total absence of PAI-1 in her blood. These observations have provided important insights into the true function of PAI-1 and the fibrinolytic system in the body. Further studies will provide valuable information for the future treatment of this and other patients with PAI-1 deficiency and for studying how PAI-1 actually works inside the body. We are currently attempting to disrupt the PAI-1 gene in mice to create an animal model for this human disease.

Bone Marrow Transplantation

Bone marrow transplantation is being used with increasing frequency to treat a variety of diseases, including several types of leukemia and a number of other cancers. In this procedure, doses of radiation and chemotherapy are given that are designed to destroy the patient's diseased bone marrow. As this would ordinarily prove fatal, the patient is then "rescued" by the transplantation of marrow from a healthy donor. Our laboratory has had a long-standing interest in bone marrow transplantation and has developed techniques to monitor what happens to the blood—to the patient's own blood cells and those received from the donor—following the procedure.

Currently the major obstacle to the more widespread use of bone marrow transplantation is an often-fatal complication called graft-versus-host disease (GVHD). This is caused by small differences between the genes of patient and donor that cannot yet be detected. We have begun a new research effort to identify the gene or genes responsible for GVHD. If this effort is successful, the process of matching patients and donors will be much improved and bone marrow transplantation may become considerably safer and more widely applied.

Uncovering the Molecular Basis of X-linked Disorders

Jane M. Gitschier, Ph.D.—Assistant Investigator

Dr. Gitschier is also Associate Professor of Medicine (Genetics) at the University of California, San Francisco. She received a B.S. degree in engineering science from Pennsylvania State University, an M.S. in applied physics from Harvard University, and a Ph.D. in biology from the Massachusetts Institute of Technology. She did postdoctoral research with Richard Lawn at Genentech, before joining the faculty at the University of California.

X-LINKED disorders result from mutations in genes on the X chromosome, one of the two sex chromosomes in humans. These common inherited conditions almost exclusively affect males because they lack the protection of a second, normal X chromosome. Females, though they have such protection, can be silent carriers of X-linked diseases and genetically transmit them to their sons. Consequently, X-linked diseases are often referred to as "sex-linked."

Our laboratory is interested in uncovering the molecular biology of some of these disorders. We are seeking to identify both the genes responsible for several diseases of unknown biochemical basis and the underlying mutations.

Much of our effort is concentrated on a particular region of the X chromosome called Xq28, the terminal portion of the chromosome's long arm. To date, 24 inherited disorders have been mapped to this region, which consists of approximately 9 million base pairs of DNA. As this density of disease loci is unusually high compared with that found in other regions of the chromosome, it suggests that Xq28 may be very rich in genes.

The incidence of Xq28-linked disorders is variable. Color blindness is extremely common, affecting about 1 of every 10 males. However, the majority of diseases occur infrequently, making accurate genetic mapping difficult. These rare diseases are clinically diverse and of unknown biochemical basis. They include Emery-Dreifuss muscular dystrophy, nephrogenic diabetes insipidus, adrenoleukodystrophy, incontinentia pigmenti, and dyskeratosis congenita.

We are attempting to isolate genes in Xq28 and determine whether they are mutated in individuals affected by Xq28-linked disorders. To date we have isolated six candidate genes and have collected 51 patient samples, including at least one example of most of the rare diseases. In some cases the sequence of the cloned gene has provided a clue to its function and suggested possible involvement in a particular disease. On the other hand, one of the isolated genes does not appear to produce a protein. Until a match is made between candidate gene and inherited disease, we are screening all patient samples for mutations in all genes, searching for gross rearrangements in genes as well as single-base alterations.

One Xq28-linked disorder we have studied in depth is classic hemophilia, or hemophilia A. It results from mutations in the gene coding for a blood coagulation protein called factor VIII. The gene was identified previously because part of the amino acid sequence was known. Our laboratory is interested in understanding what types of mutations lead to hemophilia A and how these are generated.

Hemophilia is well suited to studies on mutagenesis because it is clinically heterogeneous, implicating a wide variety of mutations in the factor VIII gene. Correlating the types of mutations with particular clinical findings may be very helpful in understanding the role of factor VIII in coagulation. To that end we recently completed a study of mutations that lead to amino acid changes in the protein. These mutations, which affect a single base pair in DNA, were revealed by a sensitive technique called denaturing gradient gel electrophoresis. By comparing the amino acid sequence of factor VIII with sequences from evolutionarily related proteins, we were able to infer that some mutations are likely to destroy activity by disrupting the protein's structure, while others alter amino acids that play a role unique to factor VIII function.

Hemophilia can occur in families lacking a genetic history of hemophilia, reflecting newly arising mutations. It is possible to uncover the origin of factor VIII gene mutations in these cases. In the past, mutations were assumed to occur as isolated events in either eggs or sperm. However, we have found evidence in several sporadic cases for "mosaicism"—i.e., a mixture of cells with and without mutation within the individual. These results suggest that mutations can occur during embryological development. In addition, these findings demonstrate that the unaffected mother of a child with sporadic disease may be at substantial risk of having a second child with the disease.

Another aspect of the hemophilia-related research has been a continued investigation of the factor VIII gene. We discovered that it may en-

code two additional proteins. The two extra messenger RNAs, termed factor VIII–associated transcripts A and B, emanate from an intron, one of the noncoding regions of the factor VIII gene. The function of the gene is unknown, but the RNAs are produced at high levels in many different cell types, suggesting a "housekeeping" role. It is interesting to speculate that the A and B transcripts influence the expression of factor VIII and that mutations in them may in turn be responsible for some instances of hemophilia.

Finally, we are attempting to create a hemophilic mouse. Genes on the X chromosome in humans are located on the X in other mammals as well, and hemophilia has been recognized in male sheep, dogs, and cats. It would be preferable, however, to use a mouse model for rapid testing of factor VIII products and for future gene therapy experiments. In progress are experiments to disrupt the normal mouse factor VIII gene, in hopes that male mice bearing the disruption will have hemophilia. If successful, this approach will be extended to create mouse models for other Xq28-linked disorders.

Membrane Lipids and Cell Regulation

John A. Glomset, M.D.—Investigator

Dr. Glomset is also Professor of Medicine and Biochemistry at the University of Washington School of Medicine. He received his M.D. and M.D./Ph.D. degrees from the University of Uppsala, Sweden. He then joined the Department of Medicine at the University of Washington. He received an honorary M.D. degree from the University of Oslo for his discovery of a plasma enzyme, lecithin:cholesterol acyltransferase (LCAT). Dr. Glomset is a member of the National Academy of Sciences.

ONE of the major research areas in my laboratory concerns the phosphorus-containing lipids (phospholipids) of animal cell membranes. In particular, we have been focusing attention on the biosynthesis and function of phospholipids that contain a polyunsaturated fatty acid, arachidonic acid. This fatty acid or its precursor, linoleic acid, must be present in the diet if animals are to reproduce, develop, grow, and maintain their health. The basis for this requirement is not fully understood, but most of the arachidonic acid that becomes available to animal cells is rapidly incorporated into membrane phospholipids. Furthermore, in response to various stimuli, a small amount of the arachidonic acid is released from the phospholipids and converted into products that can trigger many different cell functions.

The mechanisms that promote the incorporation of arachidonic acid into cell membrane phospholipids remain to be characterized, but it is clear that the metabolic pathways involved are not the classical ones described in most textbooks. For example, phosphatidylinositol (PI), a phospholipid that plays an important role in cell signaling, typically contains high amounts of arachidonic acid. As much as 80 percent of the PI present in animal cell membranes consists of a molecular species that contains arachidonic acid and a saturated fatty acid, called stearic acid. Experiments with radioactive precursors have shown that cells form this exceptional species of PI, *sn*-1-stearoyl-2-arachidonoyl (SA) PI, several hours after forming PIs that contain other fatty acids.

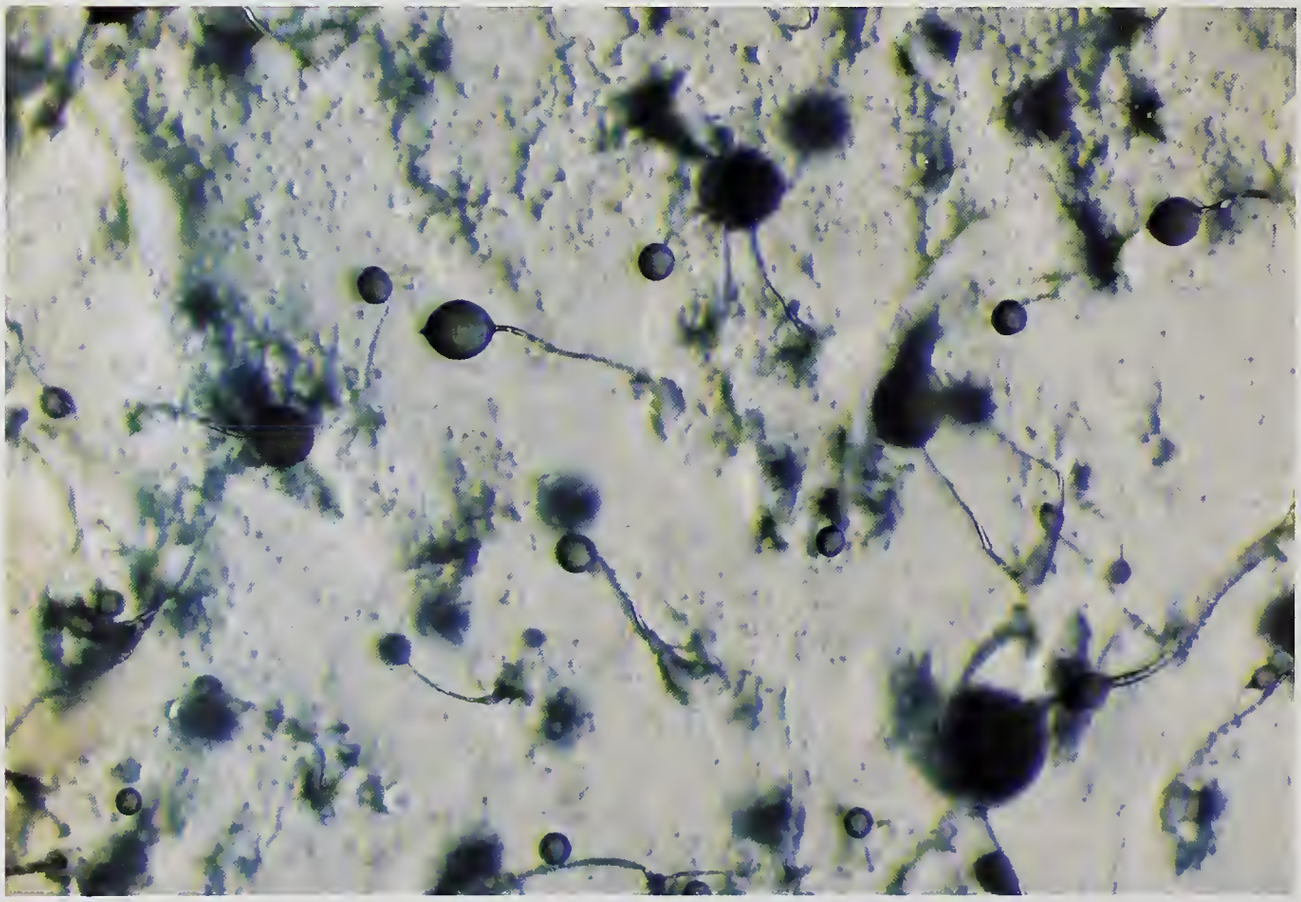
To investigate the special metabolic pathway that forms SA PI, we recently examined the possibility that animal cells in culture might be able to incorporate a radioactive precursor, *sn*-2-arachidonoyl monoacylglycerol, into this phospholipid. Our experiments showed that to be the case. Furthermore, follow-up incubation experiments with cell fractions identified two new enzyme activities that seemed to be involved. A monoacylglycerol kinase activity could catalyze the phosphorylation of *sn*-2-arachidonoyl monoacylglycerol to form the phospholipid *sn*-2-arachidonoyl lysophosphatidic acid, and a stearoyl transacylase could convert the lysophosphatidic acid into SA phosphatidic acid. In the presence of appropriate cofactors, other enzymes could convert the phosphatidic acid into the corresponding species of PI. Thus it appeared that we had discovered a new pathway of PI biosynthesis.

More recent experiments have suggested that additional steps may contribute to the pathway. We have been able to solubilize the stearoyl transacylase from membranes and examine its specificity. Surprisingly, the enzyme can use SA PI as a stearoyl donor in the transacylase reaction. When it does, a major product of the reaction is *sn*-2-arachidonoyl lysophosphatidylinositol. If the enzyme forms this product in intact cells, the lysophosphatidylinositol might well have some special function or metabolic fate. To investigate this possibility we are currently conducting a search for enzymes that catalyze the conversion of lysophosphatidylinositol into PI. If we find some, we will characterize their activities in order to define the pathway more completely. Then we will attempt to localize the various enzymes within cells in order to explore the pathway's intracellular role.

One potential role of pathways that form arachidonic acid–containing phospholipids may relate to the fine structure of animal cell membranes. Thus the arachidonic acid in phospholipids might conceivably affect their ability to pack tightly with one another in membranes. We began to investigate this possibility some time ago in parallel with our studies of phospholipid biosynthesis. We used a computer-based approach to examine the effects of arachidonic acid on the conformation and packing properties of model phospholipids in simulated monolayers. The results of this molecular modeling approach have suggested that arachidonic acid–containing phospholipids may be able to form straight conformations and pack together in regular, tight arrays.

We are currently attempting to test this possibility, in collaboration with Howard Brockman of

the Hormel Institute. If we find that semisynthetic arachidonic acid–containing phospholipids form tightly packed arrays in experimentally produced phospholipid monolayers and bilayers, we will expand our experiments to include animal cell membranes. A demonstration of such tightly packed arrays in cell membranes would raise fundamental questions about the role of phospholipid domains in programming membrane-associated events.



Fruiting bodies of the slime mold Dictyostelium, *each a mass of spore cells supported by a thin column of stalk cells.*

Research of Richard Gomer.

Determination and Maintenance of Cell Type

Richard H. Gomer, Ph.D.—Assistant Investigator

Dr. Gomer is also Assistant Professor of Biochemistry and Cell Biology at Rice University and Adjunct Assistant Professor of Cell Biology at Baylor College of Medicine. He received his B.A. degree in physics from Pomona College and his Ph.D. degree in biology from the California Institute of Technology. He was a postdoctoral fellow in the laboratory of Richard Firtel at the University of California, San Diego.

OUR laboratory is interested in the general problem of differentiation and morphogenesis. We are trying to understand at a molecular level some of the factors that determine the cell type into which an embryonic cell differentiates and how the ratios of the different cell types are then maintained in an organism. As a model system, we are using the slime mold *Dictyostelium discoideum*.

Dictyostelium normally exists as undifferentiated single cells called amoebae that eat bacteria in soil and decaying leaves and proliferate by cell division. When the amoebae eventually overgrow their food supply and starve, they aggregate together in groups of about 100,000. Roughly 80 percent of these cells become spores. (A spore is a cell with a tough outer coat that forms an "escape capsule.") The remaining 20 percent of the cells form a stalk about 2 mm high that holds the spore cells off the ground. A spore, dispersed by the wind, will crack open to release an amoeba, which may luckily find itself in the midst of a new supply of bacteria. The advantage of this organism is its simplicity: cells differentiate into just two main types.

Determining Cell Fate

In the presence of a food source, *Dictyostelium* cells grow to a certain size and then separate their chromosomes to opposite sides of the cell and divide in two. The cycle of growth and division then repeats. In a field of cells, there will be cells at all different phases of this cycle.

Dictyostelium uses a simple and elegant mechanism based on this cycle to determine whether a cell will become stalk or spore. When the cells starve, those cells that have recently separated their chromosomes and divided will become stalk cells; the remaining cells will become spores. As long as the cells are randomly distributed with respect to the phase of their cell cycles, there will always be the proper percentage of cells in the "stalk" quadrant. We refer to this as a musical chairs mechanism, since the decision of any given cell to become either stalk or spore is made by the phase of the cycle that the cell happens to be in at the time of the differentiation signal (starvation).

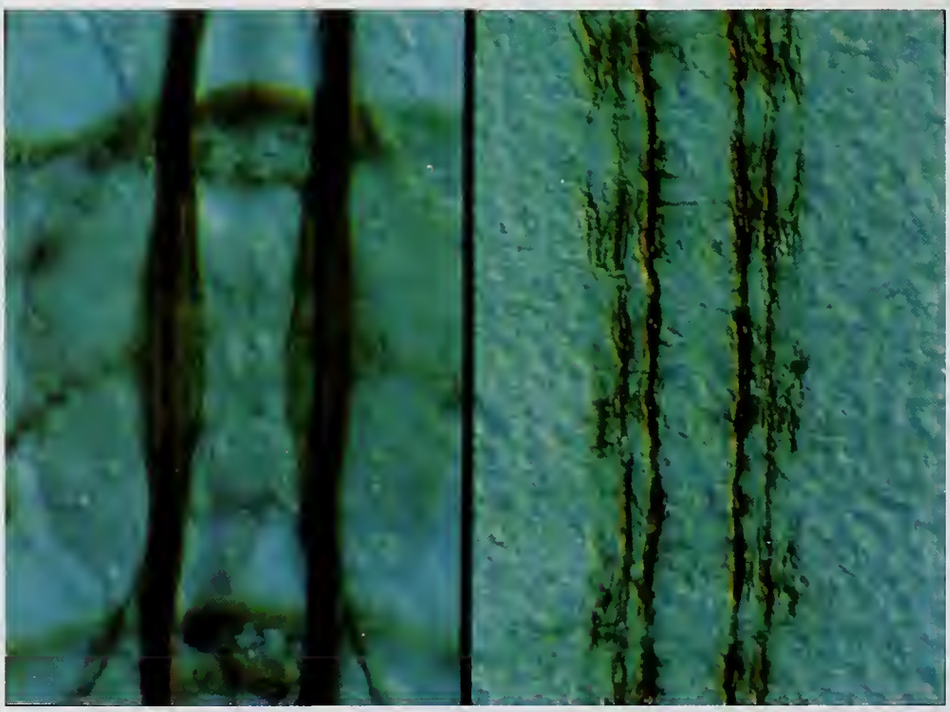
Cell-type choice determination mechanisms of this sort may operate in humans, and aberrations could thus lead to birth defects. We are developing a variety of techniques that use DNA inserted into *Dictyostelium* cells to identify genes involved in cell-type differentiation. In the past year we have engineered DNA constructs to mutate *Dictyostelium* cells three ways: by having too many copies of a sequence of DNA, by inserting foreign DNA into the chromosomal DNA, and by making RNA that binds to and renders useless the RNA encoding a specific protein. In preliminary experiments with these constructs, we have identified mutants that may have abnormal ratios of the two cell types. In addition, to examine the extent of the linkage between the cell cycle and cell-type choice, we have treated cells with drugs that disrupt the cell cycle. These experiments confirmed that the cell-type choice mechanism is linked to the cell cycle and does not use a separate timer.

Sensing Cell Density

During *Dictyostelium* development, cells turn on the stalk- or spore-specific genes only when above a certain cell density. Being able to sense whether other cells are nearby represents a paradigm for possible mechanisms that would allow, for instance, liver cells or others to sense how much of the body is composed of liver cells. At present, little is known about the molecular mechanisms whereby the size and density of a tissue are sensed by individual cells. Such mechanisms would be centrally involved in the regulation of growth during development, wound healing, and regeneration. In addition, an important and relevant aspect of tumor cells is that they have lost their ability to regulate the size and/or density of the tissue and, as a result, proliferate. One way this could happen would be if the tumor cells could no longer properly sense the total mass of the tissue.

In experiments funded by the National Institutes of Health, we have found that *Dictyostelium* cells sense whether they are near other starving cells by, upon starvation, secreting small quantities of a protein that we have named density-sensing factor (DSF). Cells are sensitive to DSF: above a threshold concentration they will express cell-type-specific genes. We have done theoretical diffusion calculations and have found, in agreement with our observations, that DSF secreted from cells that are quite far from other DSF-secreting cells diffuses away so quickly that it never reaches the threshold concentration. However, at a sufficiently high density of starving cells, the DSF concentration will be above the threshold value. We have purified DSF protein and have found that DSF eventually breaks down to small protein fragments that are much more effective in activating differentiation. This might be a safety mechanism to allow cells that cannot find other starving cells to stimulate themselves and perhaps form an isolated spore. Interestingly, DSF is made and stored in growing cells and is only secreted upon starvation.

We have cloned and sequenced the DNA encoding DSF. This has been used in turn to obtain the sequence of amino acids for the DSF protein. Computer comparison with data banks of other protein sequences indicates that DSF is not closely related to any known protein. We have used the DSF DNA to generate *Dictyostelium* cells that do not make DSF; these cells do not aggregate unless DSF protein is added to their medium. *Dictyostelium* cells, which eat bacteria that are almost their size, starve asynchronously. The behavior of the cells lacking DSF suggests that its function is to coordinate development by triggering aggregation when most of the cells in a given area have starved, as signaled by DSF secretion, and are ready to form one large fruiting body. Finally, we have made antibodies that will allow us to examine where DSF is stored in *Dictyostelium* cells and to search for similar proteins in other organisms.



These two photographs show the expression of fasciclin II (brown axon bundles) in the developing central nervous system of the grasshopper (left) and Drosophila *embryos (right). Although separated by well over 300 million years, these two insects continue to express the protein on a specific subset of longitudinal axon pathways. The brown staining results from HRP immunocytochemistry using anti-fasciclin II antibodies in both species.*

Left photograph from Allan Harrelson and Corey Goodman. Right photograph from Gabriele Grenninglob and Corey Goodman.

Growth Cone Guidance and Neuronal Recognition in Drosophila

Corey S. Goodman, Ph.D.—Investigator

Dr. Goodman is also Class of '33 Professor of Neurobiology and Genetics in the Department of Molecular and Cell Biology at the University of California, Berkeley, and Adjunct Professor in the Department of Physiology at the University of California, San Francisco. He received his B.S. degree in biology from Stanford University and his Ph.D. degree in developmental neurobiology from the University of California, Berkeley. His postdoctoral work in developmental neurobiology was done at the University of California, San Diego. Prior to his present position, Dr. Goodman was a faculty member at Stanford University. His honors include the Alan T. Waterman Award from the National Science Board.

WE are interested in understanding the molecular mechanisms by which neuronal growth cones find their way toward, and ultimately recognize, their correct targets during development. Growth cones navigate over long distances and often through a series of complex choice points, appearing to follow signals on the surfaces of cells and in the extracellular environment. Our studies are aimed at uncovering the molecules and mechanisms that impart specificity to the developing nervous system and thus allow growth cones to recognize their correct pathways and targets. To address these issues, we use molecular genetic approaches in the fruit fly *Drosophila*.

Guidance of Pioneers

Within and just outside of the developing central nervous system (CNS), certain glial cells and other special midline cells provide instructive information for the differential guidance of the initial, "pioneering" growth cones as they choose which cells to extend toward or along. For example, a specific subset of cells at the midline appears to provide an attractive cue for the growth cones that extend toward the midline and pioneer the commissural axon pathways that connect the two sides of the developing CNS. Similarly, a specific pattern of longitudinal glia appears to provide an important substrate for the formation of the longitudinal axon tracts that connect adjoining segments of the CNS. We have conducted a large-scale screen for mutations that perturb the guidance of pioneering growth cones.

Of the hundreds of new mutations identified in this screen, mutations in three genes are of particular current interest. Mutations in *longitudinals lacking* (*lola*) have a dramatic phenotype: although both commissural and peripheral pathways are normal, as are most other aspects of embryogenesis, the CNS of these mutants lack all longitudinal axon tracts. In *lola* mutant embryos, the longitudinal glia are born, initially migrate, and divide as normal; the earliest defect is seen about the time that the first growth cones contact these glia and fail to extend along them.

Mutations in the second gene, *commissureless*, have an equally dramatic phenotype: although all other axon pathways are normal, the CNS of these mutants lack all commissural pathways. They also have normal peripheral nervous system (PNS) axon pathways, muscles, sensory organs, and body organization. In these mutant embryos, the growth cones of CNS neurons do not extend across the midline and commissures never form, although other aspects of embryonic pattern formation and neuronal development appear normal. The *commissureless* gene product is a good candidate to be either the signal or the receptor for the guidance of growth cones toward the midline.

Mutations in a third gene, *roundabout* (*robo*), lead to a dramatic misrouting of the growth cones that normally pioneer the MP1 pathway. For example, the MP1 growth cone extends across the anterior commissure where it contacts its homologue from the other side, rather than proceeding posteriorly. In contrast, many other longitudinal pathways develop as normal in *robo* mutant embryos.

Pathway Recognition

Once the initial axon pathways are established, the predominant guide for "follower" growth cones is the surface of the earlier axons in these pathways. Growth cones are able to distinguish one particular axon bundle, or fascicle, out of an array of many. The experimental analysis of these phenomena led to the labeled pathways hypothesis, which holds that axon pathways are differentially labeled by recognition molecules that enable growth cones to navigate through complex choice points. To identify such recognition molecules, we used an immunological approach to identify and subsequently clone the genes encoding five different surface glycoproteins: fasciclin I, fasciclin II, fasciclin III, fasciclin IV, and neuroglian (a sixth, connectin, has recently been cloned using a different method, as described in

the following section). Five of the six glycoproteins are dynamically expressed on overlapping subsets of growth cones, axon fascicles, and glia during embryonic development; the sixth, neuroglian, is more broadly expressed on the surface of most axons and glia. Fasciclin II, neuroglian, and fasciclin III are members of the immunoglobulin superfamily; connectin is a member of the leucine-rich repeat family; fasciclin I and fasciclin IV are neither related to each other nor thus far to anything else in the data bank. Five of the six proteins (except for fasciclin IV) can function as homophilic cell adhesion molecules.

Genetic analysis has shown that fasciclin II is indeed a neuronal recognition molecule that plays an important role in specific growth cone guidance. The fasciclin II protein is normally expressed on a subset of growth cones and axons that pioneer and selectively fasciculate in the MP1 axon fascicle; later it is expressed on several other axon pathways. In embryos that are mutant for the *fas II* gene, although these specific growth cones extend, they do not properly recognize one another and fail to fasciculate. These studies are the first molecular confirmation of the existence of functional labels on specific axon pathways in the developing nervous system.

A major current goal is to determine the molecular machinery whereby the activation of fasciclin II, by either another fasciclin II molecule or a different ligand on the surface of neighboring neurons, instructs a fasciclin II–expressing growth cone to extend toward and along these specific cells. We have established a highly sensitive assay that utilizes a mutation in the *fas II* gene that reduces the amount of fasciclin II protein to the minimal amount required for *fas II* function. Using this assay we are looking for other genes in which a 50 percent reduction in the level of their protein product results in a failure of *fas II* function. In this way we are in the process of identifying genes that appear to encode products that interact with the *fas II* gene and thus function in the events of neuronal recognition.

Our previous studies have shown that the expression of surface recognition molecules is dynamic and regional on the surface of individual neurons; i.e., parts of the cell are differentially labeled in accordance with the processes around it for which it has a selective affinity. We wish to uncover the signals that instruct a neuron, as it navigates from one pathway to another, to change the expression of surface recognition molecules on its growth cone. For example, the growth cones of many commissural interneurons extend right past longitudinal axon pathways on their own side; once they cross the midline, they display a high affinity for one of these axon pathways on the other side of the CNS. This change in the behavior of the growth cone is accompanied by an equally dramatic change in its expression of surface recognition molecules; neurons express different molecules on the surface of their longitudinal and commissural processes. We are investigating whether extension across the midline is required for growth cones to make the appropriate changes in surface expression and subsequent pathway choices, by examining the behavior of specific growth cones in *commissureless* mutant embryos in which guidance toward the midline does not occur. Thus far it appears that some aspects of longitudinal guidance are perturbed in these mutant embryos.

Target Recognition

Having navigated along a series of pathways, growth cones are ultimately capable of recognizing their correct target cells. In the *Drosophila* embryo, the specificity of neuronal growth cones for target cells is most clearly studied in the ability of motoneuron growth cones to recognize specific muscle fibers. A large-scale molecular genetic screen has led to the identification of several genes that are expressed by different subsets of undifferentiated muscle fibers prior to innervation. A new gene, *connectin*, and a previously identified gene, *Toll*, encode membrane proteins that are members of the leucine-rich repeat family of cell adhesion and signaling molecules. The *connectin* gene encodes a protein that is expressed on the surface of eight (of the 31) muscle fibers in each abdominal segment and on the growth cones and peripheral axons of the very motoneurons that innervate these eight muscle fibers. The growth cones of this subset of motoneurons, and the undifferentiated muscle fibers they innervate, express the connectin protein prior to contact and synapse formation. *In vitro* studies show that the connectin protein is capable of functioning as a homophilic cell adhesion molecule. Thus this protein is a good candidate to be a bona fide target recognition molecule involved in the formation of specific synaptic connections. A genetic analysis of *connectin* function is presently in progress. In addition, a large-scale genetic screen is under way to identify mutations that perturb the guidance of motoneuron growth cones.

Mechanisms of Immunological Self-Tolerance and Autoimmunity

Christopher C. Goodnow, B.V.Sc., Ph.D.—Assistant Investigator
Dr. Goodnow is also Assistant Professor of Microbiology and Immunology at Stanford University School of Medicine. He was educated in the United States and Australia and received B.S. and veterinary degrees from the University of Sydney. After training in molecular immunology with Mark Davis at Stanford University, he returned to the University of Sydney to complete doctoral and postdoctoral studies on immunological tolerance in the laboratory of Antony Basten.

EACH individual B and T lymphocyte in the immune system expresses antigen receptors of one type on its surface, which confer on the cell an ability to recognize one of the millions of different antigens; and there are millions of different lymphocytes in the immune system. Given the system's annihilative powers, it is remarkable that tissue components of our own bodies are spared during immunological attacks on invading foreign organisms. Normally the immune system can recognize one's tissue components as "self" and tolerate them. Self-tolerance is lost, however, in a variety of "autoimmune" diseases, such as systemic lupus erythematosus, type I diabetes mellitus, and rheumatoid arthritis, resulting in inexorable destruction of particular organs and tissues. The mechanisms that maintain self-tolerance in healthy individuals, and the factors that lead to its breakdown in autoimmune disease, are the main focus of our laboratory.

It has been theorized for many decades that self-tolerance might somehow result from the silencing or elimination of lymphocytes bearing antigen receptors that happen to recognize self antigens. To determine whether this idea was correct, however, has been difficult, since it is almost impossible to track the life of any one cell among millions.

Advances in biotechnology, in particular the advent of transgenic mice, have opened doors to the development of ways to follow the life of particular immune cells *in vivo*. Transgenic mice are genetically altered at the outset of embryonic development by microinjecting carefully designed cassettes (transgenes) into fertilized oocytes. With colleagues in the laboratory of Antony Basten at the University of Sydney, we produced transgenic mice in which most of the B lymphocytes expressed identical, rather than widely differing, antigen receptors.

This was done by introducing transgenes that coded for a single antibody molecule (since antibody molecules serve as antigen receptors on B lymphocytes). The particular molecule was one that recognized and bound a foreign protein, hen egg-white lysozyme, and because the transgene was expressed in essentially all the B lymphocytes, they all now recognized lysozyme in an identical fashion.

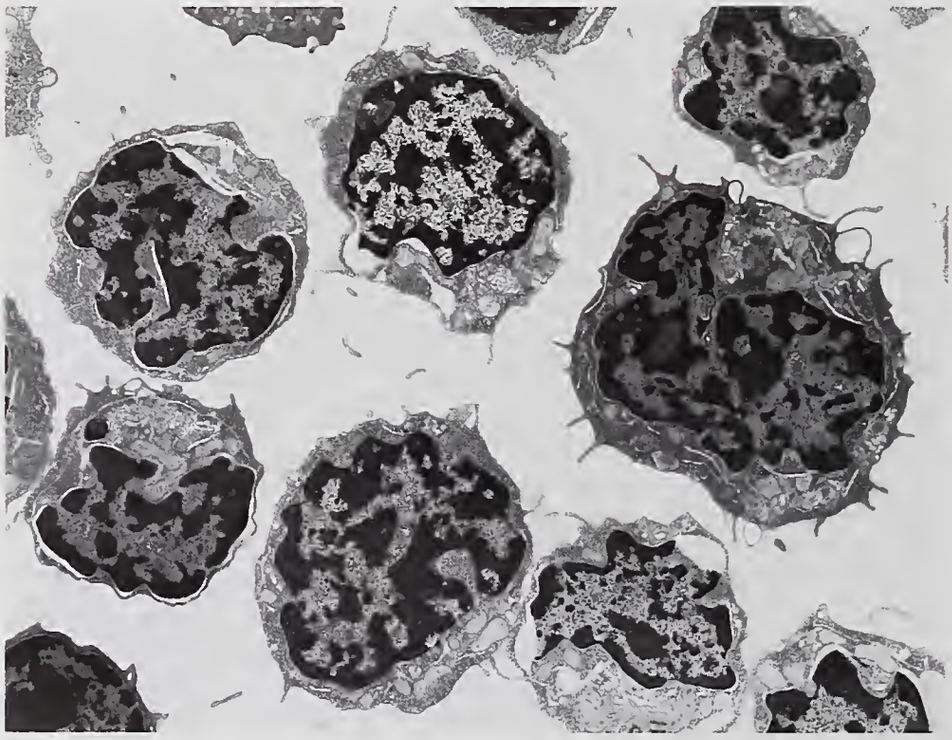
The extraordinary abundance of lysozyme-binding B cells in the transgenic mice has made it possible to track the development and fate of these cells in the body. To determine the fate of B cells that might happen to recognize a self antigen rather than a foreign one, we prepared additional transgenic mice in which transgene cassettes led to the synthesis and production of lysozyme by the mouse itself. When the two types of mice are mated, a fraction of their offspring inherit both types of transgene and thus contain large numbers of lysozyme-binding B cells that encounter lysozyme expressed as if it were a normal self constituent.

The self-reactive B cells that develop in these mice undergo one of three distinct fates. First, if lysozyme is present at too low a concentration or it binds to the B cell's receptors with too low an affinity, no signal is registered by the cell. As a result, the B cells are neither activated nor rendered tolerant but remain fully capable of mounting an antibody response if they receive a stronger stimulus subsequently. Second, when higher concentrations of soluble lysozyme are bound with sufficient affinity, the B cells continue to follow their normal maturation program but store some sort of negative signal, such that they are unable to mount an antibody response when stimulated to do so. Third, if lysozyme is encountered in a form where it is displayed on cell surfaces, and thus bound by the B cell with extremely high avidity, developing B cells in the bone marrow are triggered to abort their maturation program and die.

These distinct cellular events help to explain the long-standing observation that self-tolerance is an actively acquired but incomplete process and suggest a number of possible avenues along which autoimmunity may ultimately develop. For example, our laboratory has found that tolerant B cells that have stored a negative signal can nevertheless recover the ability to mount an efficient antibody response. Recovery of function

and breakage of tolerance depends, however, on a balance between stimuli that promote antibody formation and stimuli that reinforce tolerance. One line of research in our laboratory involves developing transgenic mouse models for studying how the interplay between these factors shapes the development and remission of autoimmune disease *in vivo*. This research is funded by the National Institutes of Health.

Identifying the molecular events that prevent tolerant cells from mounting an antibody response, or lead to the death of the autoreactive cells, will be crucial for fully understanding how self-tolerance is induced and maintained. Our approach to this problem is severalfold. First, we are continuing to delineate the precise conditions under which tolerance rather than activation is induced in B cells *in vivo*. Second, we have developed tissue culture systems that reproduce the cellular responses observed in the whole animal, allowing much closer scrutiny of the sequence of events accompanying tolerance induction. In autoreactive B cells that are unable to mount an antibody response, this is due to interference with the cell's capacity to replicate and to differentiate into an antibody-secreting cell. For B cells that are triggered to die, the primary event appears to be an arrest of cell development at an immature, short-lived stage of development. Third, we are using a combination of biochemical, molecular, and genetic strategies to search for the key differences in biochemical signaling cascades or changes in gene expression that underlie these events.



Self-reactive B lymphocytes sorted from the bone marrow and visualized by electron microscopy. The antigen receptors on these cells have already been triggered, and as a result the cells are destined to die within a few days in vivo. *At this stage, however, the process is reversible and the cells show none of the ultrastructural changes that precede programmed cell death.*

Research of Suzanne Hartley and Christopher Goodnow.

Developmental Control of Gene Expression

Rudolf Grosschedl, Ph.D.—Assistant Investigator

Dr. Grosschedl is also Associate Professor of Microbiology and Immunology and of Biochemistry and Biophysics at the University of California, San Francisco. He completed his undergraduate studies on the replication of lambdoid bacteriophages in the laboratory of Gerd Hobom in Freiburg, West Germany. His graduate studies, on the regulation of histone gene expression, were carried out in the laboratory of Max Birnstiel in Zurich, Switzerland. Dr. Grosschedl spent his postdoctoral years with David Baltimore at the Massachusetts Institute of Technology and the Whitehead Institute.

THE process of terminal differentiation turns a multipotential cell into a cell that carries out a particular function or synthesizes a specific product. The lymphoid B cell lineage ultimately generates a cell that secretes antibody. During cell differentiation, genes that encode the antibody or associated proteins are expressed in a defined cell type–specific and temporally ordered pattern.

Transcription of the μ immunoglobulin (Ig) gene encoding the heavy chain of the antibody can be detected in virtually all lymphocytes. By contrast, the κ immunoglobulin light-chain gene is transcribed only in late-stage B cells, and the *mb-1* gene encoding an antibody-associated protein is expressed only in early-stage B cells. The goal of our research is to gain some insight into the molecular mechanisms that mediate the developmental control of lymphoid-specific gene expression.

Tissue-Specific Regulation of Ig Gene Expression in a Transgenic Model

We are attempting to understand how multiple cis-acting regulatory information of a gene is integrated to govern the correct developmental pattern of expression. Toward this goal, we transferred rearranged wild-type and mutated Ig μ genes into the mouse germline and analyzed expression in various tissues and developmental stages. Our first set of experiments suggested that lymphoid-specific expression is dependent upon multiple tiers of regulation.

We found that the intragenic enhancer directs high-level expression in lymphoid tissues and basal expression in most but not all nonlymphoid tissues. Basal expression appears to be governed by negative regulation, since mutations of the μE5- and μE2-binding sites in the enhancer increased μ gene expression in many nonlymphoid tissues to levels similar to those found in lymphoid tissues. However, the "off-state" of expression observed in a few other nonlymphoid tissues was not affected by these mutations, suggesting that another tier of negative regulation may be involved in further decreasing basal expression.

An additional set of experiments suggested that the Oct-binding site in the promoter contributes to the positive regulation of μ gene expression during B cell differentiation. A mutation in the Oct-binding site decreased μ gene expression in fetal pre–B cells (which represent early B cells) by a factor of 5, but decreased μ gene expression in stimulated splenocytes (representing late-stage B cells) by a factor of 100 to 200. Together these experiments indicate that the multiple modes of regulation are necessary to maximize the difference between the on-state and off-state of gene expression *in vivo*.

In a second set of experiments, we addressed the regulation of the accessibility of binding sites for nuclear factors in native chromatin. With the aim of uncoupling changes in chromatin structure from the process of transcription, we linked DNA fragments comprising the μ enhancer region to the promoter for bacteriophage T7 RNA polymerase and incorporated the gene construct into the mouse germline. Subsequently, we examined the accessibility of the T7 promoter in isolated pre–B cell nuclei by adding exogenous T7 RNA polymerase and measuring the synthesis of T7-specific RNA.

The T7 promoter was accessible in 7/7 lines when linked to the μ enhancer. By contrast, the T7 promoter alone was only weakly accessible in two lines and not accessible in six. Because the enhancer fragment used in this experiment lacked any known promoter activity, we are currently using this assay to examine which sequences confer accessibility in native chromatin.

Novel Lymphoid-Specific Regulators of Gene Expression

To identify and clone novel lineage-specific transcriptional regulators, we adopted two approaches. First, we included in our analysis of lymphoid-specific gene expression the *mb-1* promoter, which has a cell type–specific pattern of activity distinct from that of the Ig promoters. We found that the *mb-1* promoter is functional in

pre-B and surface Ig-positive B cells but not in later-stage B cells or T cells. In our study of the *mb-1* promoter, we identified a novel nuclear factor, termed early B cell factor (EBF).

EBF was shown to bind a functionally important sequence in the distal promoter region. The factor was found to be present specifically in pre-B and surface Ig-positive B cells, thus paralleling the pattern of activity of the *mb-1* promoter. Recently we purified EBF to homogeneity and obtained partial amino acid sequences. Based on this information, we isolated cDNA clones that correspond to EBF. Currently we are characterizing its DNA-binding and regulatory properties.

The second approach to identify lineage-specific regulators of gene expression consisted in differential screening of a pre–B cell (minus erythroid) cDNA library. One of the cDNA clones was found to encode a DNA-binding protein with homology to the chromosomal nonhistone high-mobility group protein HMG-1 and to regulators of cell specialization. This pre-B and T cell–specific DNA-binding protein, termed LEF-1 (lymphoid enhancer-binding factor 1), was found to bind a functionally important site in the T cell receptor (TCR) α gene enhancer.

LEF-1 was shown to participate in the regulation of the TCRα enhancer. First, co-transfection of cells lacking endogenous LEF-1 with TCRα enhancer-containing reporter genes together with an LEF-1 cDNA expression plasmid increased enhancer function. Second, in the context of the intact minimal TCRα enhancer, the LEF-1–binding site is required for full enhancer function.

We showed that the HMG domain of LEF-1 encompasses a sequence-specific DNA-binding domain that interacts with the minor groove of DNA. Moreover, we defined amino acids that are conserved among various members of the family of HMG-domain proteins as residues important for DNA binding.

In previous experiments, we found that LEF-1 by itself is unable to augment basal promoter activity. This observation raised the possibility that LEF-1 functions by aiding the binding and/or action of other nuclear factors. Recently we demonstrated that DNA binding by the HMG domain induces a sharp bend in the DNA helix. Therefore, we examined whether LEF-1–induced DNA bends can facilitate communication between proteins bound at distant sites.

We replaced one binding site for the bacterial integration host factor with one for LEF-1 in the *att*P locus of bacteriophage λ and found that LEF-1 is capable of precisely aligning widely separated integrase (Int) protein-binding sites and stimulating Int-dependent recombination. These data suggest that LEF-1 can serve as an "architectural" element in the assembly of higher-order nucleoprotein structures. Currently we are examining the role of DNA bending for the regulation of TCRα enhancer function and are attempting to gain insight into the importance of LEF-1 for lymphoid differentiation by targeted gene inactivation.

Polypeptide Hormone Gene Regulation

Joel F. Habener, M.D.—Investigator

Dr. Habener is also Professor of Medicine at Harvard Medical School and Associate Physician and Chief of the Laboratory of Molecular Endocrinology in the Department of Medicine at Massachusetts General Hospital, Boston. He obtained his B.S. degree in chemistry at the University of Redlands and his M.D. degree at the University of California, Los Angeles. After medical residency training at the Johns Hopkins Hospital, he spent two years in research at the National Cancer Institute. Dr. Habener completed his medical training in endocrinology and metabolism at Massachusetts General Hospital.

OUR laboratory seeks an understanding of the molecular processes involved in the regulation of gene expression. The general hypothesis being tested is that peptide hormones are important regulatory molecules in conveying information among cells via ligand-receptor interactions and corresponding signal transduction, resulting in the expression of specific genes. These processes are important in determining cellular metabolic responses such as secretory activity, cellular differentiation, and growth.

Peptides in Cellular Metabolism

Peptides activate metabolic responses in cells by way of interactions with specific receptors on distant (endocrine), adjacent (paracrine), or the same (autocrine) cells. These ligand-receptor interactions lead to the activation of signal transduction pathways involving postulated cascades of protein phosphorylation enzymatically catalyzed by protein kinases, eventuating in the assembly of active transcriptional complexes. Under intensive investigation are two such signaling pathways mediated by cAMP-dependent protein kinase A and by phospholipid/diacylglycerol-stimulated protein kinase C. A major focus of the laboratory is to understand how specific phosphorylated DNA-binding proteins interact with cognate DNA sequences and thereby induce gene expression.

Genes Encoding Polypeptide Hormones

We have determined the structures, organization, and regulation of the expression of genes encoding several of the polypeptide hormones. Our work has centered on the genes encoding the glucagon and glucagon-related peptides, somatostatin, insulin, angiotensinogen, and the gonadotropins. These studies have progressed through several stages: 1) cloning of the genes and elucidation of their structures, 2) determinations of the DNA enhancer and suppressor sequences responsible for the regulation of the transcriptional expression of the genes, and 3) isolation and characterization of DNA-binding proteins involved in the regulation of expression. These studies have led to the identification of cell-specific enhancer sequences within the "promoter" regions of the genes—sequences that determine in which cellular phenotype the genes are expressed and how gene transcription responds to activator substances such as cAMP and phorbol esters. They have also led to the identification of complex, cell-specific post-translational processing of protein precursors (prohormones) encoding the peptide hormones.

Recently we determined that specific nuclear proteins bind to these important DNA enhancer sequences and that the binding specificities, as well as the transactivation activities, of these proteins are regulated by their phosphorylation.

Regulation of Glucagon and Somatostatin Gene Expression

Our analyses of the regulation of the expression of the glucagon gene in pancreatic islet cell lines reveal that islet cell–specific expression resides in at least two enhancer-like sequences and that A cell expression in the islets is determined by an additional enhancer/promoter combination. The expression of the somatostatin gene is restricted to the D cells because suppressor elements prevent expression in the glucagon-producing A cells and insulin-producing B cells. Transcriptional activation of the glucagon gene is mediated through both protein kinase C and protein kinase A pathways, whereas activation of the somatostatin gene is regulated by protein kinase A and a calcium-regulated pathway.

Cloning and Structure of a cAMP-Dependent DNA-binding Protein

In studies of the somatostatin and gonadotropin genes, we have determined that their expression is stimulated via a cAMP-dependent signal transduction pathway. We have discovered that DNA-binding proteins interact with specific, short DNA sequences to generate cAMP-responsive complexes. These DNA-protein complexes that mediate either cAMP or phorbol ester control of gene transcription share certain related structures. Cooperative interactions among several

DNA-binding proteins and with adjacent target DNA sequences appear to determine cellular specificity of gene expression as well as metabolically regulated expression. Recently we have cloned several members of a family of cAMP-responsive enhancer-binding proteins (CREBs) and have discovered a domain on one of the CREBs that is phosphorylated by the cAMP-dependent protein kinase A, as well as by additional protein kinases. The phosphorylation of CREB has a profound effect of increasing the transactivation of gene transcription.

Discovery of an Insulinotropic Peptide

We have discovered a new glucagon-like peptide related in its structure to pancreatic glucagon and co-encoded with glucagon in the precursor protein of the two hormones (proglucagon). These peptides are differentially cleaved from the proglucagon, the initial translation product, in the pancreas and the intestines. One of the peptides, glucagon-like peptide-I (7–37) produced in the intestine and released in response to oral nutrients has potent insulinotropic activities. Concentrations as low as 10^{-11}–10^{-12} M stimulate insulin secretion in the perfused rat pancreas and stimulate both cAMP formation and proinsulin gene transcription and mRNA levels in insulinoma cell lines. We have determined that the glucagon-like peptide regulates insulin secretion in humans and propose that it may be involved in the pathogenesis of certain types of diabetes mellitus due to a faulty regulation. Particular emphasis is on analyses of the β-cell receptor for the glucagon-like peptide and the mechanisms operative in the stimulation of insulin gene transcription in response to the actions of the peptide.

Our goals are 1) to characterize further the genes encoding the regulatory (DNA-binding) proteins that interact with tissue-specific enhancers and to determine how cAMP-mediated expression of the peptide hormone-encoding genes is regulated in the specific cellular phenotypes and 2) to test further the new glucagon-like peptides for their potential regulatory actions within the pancreatic islets and the intestinal tract and to explore the possible role of the peptides in diabetes mellitus.

Some of these studies were also supported by funds from the National Institutes of Health.

Structural Studies of Macromolecular Assemblies

Stephen C. Harrison, Ph.D.—Investigator

Dr. Harrison is also Professor of Biochemistry and Molecular Biology at Harvard University and Research Associate in Medicine at the Children's Hospital, Boston. He received his A.B. degree in chemistry and physics from Harvard College and his Ph.D. degree in biophysics from Harvard University. Dr. Harrison is a member of the National Academy of Sciences.

HOW do regulatory proteins activate or inhibit transcription of particular genes? How do viruses leave and enter cells? How do receptors and their ligands cycle from cell surface to cell interior and back? These questions deal with molecular recognition in the determination of cell organization. They represent groups of projects in our laboratory, all of which involve elucidation of detailed atomic structures as a prerequisite for tackling functional problems.

Transcriptional Regulatory Complexes

A common characteristic of eukaryotic transcriptional regulatory elements is the presence of sites in multiple copies varying slightly in sequence, often with two or more related proteins that can bind to them. The best-understood prokaryotic paradigm is in the immunity region of temperate bacteriophages, where two proteins, repressor and Cro, bind two sets of three sites, with appropriately graded affinities. We have made an effort to understand the mechanism of this regulatory switch, by determining the structures of a series of specific protein/DNA complexes containing the Cro protein of phage 434 or the DNA-binding domain of its repressor. We are using computational approaches to link observed structural differences among these various complexes with the corresponding free energies of binding.

We are studying structural aspects of eukaryotic transcriptional regulation, initially by examining the DNA-binding domains of regulatory proteins in complex with synthetic binding sites. GAL4, the prototype of a class of such proteins, is a regulator of galactose metabolism in yeast. It binds as a dimer to the upstream activity sequences of several genes involved in galactose and melibiose catabolism. The DNA-binding region is at the amino terminus of the 881-residue polypeptide chain. We have determined the structure of a fragment containing residues 1–65, in a specific complex with DNA. A small domain containing zinc ions (residues 8–40) recognizes a conserved CCG triplet at each end of the 17–base pair binding site, through direct, major-groove contacts. A short α-helical coiled-coil element imposes twofold symmetry. A segment of extended polypeptide chain, linking the metal-binding module to the dimerization element, specifies the length of the site. The complex has a relatively open structure, which would allow another protein to bind coordinately with GAL4. Coordinate binding of two or more proteins has been shown to be an important feature of many eukaryotic control elements.

GCN4, also a yeast transcriptional regulator, represents yet another class of DNA-binding proteins. It contains a dimerization element, generally called a leucine zipper, which forms an α-helical coiled coil about 30 residues in length. This segment is preceded in the protein sequence by a positively charged region, which has little ordered structure in the free protein but which also acquires α-helical structure when it binds to DNA. We have prepared crystals of the basic region/leucine zipper fragment of GCN4, in complex with a synthetic binding site (having a so-called AP-1 sequence). The structure determination is nearly complete. Preliminary analysis shows that each chain is a continuous α-helix. The part corresponding to the basic region lies along the major groove, and contacts can be made by suitable side chains to four base pairs on either side of a central GC.

TfIIIA, which controls 5S RNA transcription in *Xenopus*, represents the so-called zinc finger class of regulatory proteins. The finger is a small, 30-residue domain, stabilized by a tightly bound zinc ion. A recombinant fragment comprising seven of the nine fingers from TfIIIA binds to a 30–base pair DNA containing an appropriate part of the total binding site, and we have crystallized this complex. The fragment also binds specifically to 5S RNA.

Understanding how these various structures recognize their DNA sites is only a beginning. The specificities of interactions between other domains of these proteins and additional compo-

nents of a transcriptional initiation complex present even more challenging puzzles for the future.

Viruses

The small, double-stranded DNA viruses SV40 (simian virus 40) and polyoma have given us a first glimpse of virus particles that package a minichromosome in one cell and deliver it to the nucleus in another. The shells of these viruses are composed of 72 pentamers of the major structural protein VP1 and 30–60 copies each of two internal proteins, VP2 and VP3. These components package the viral DNA. The VP1 polypeptide chain folds in such a way that two large β-sheets with radially directed strands form a framework, with very tight interactions between adjacent subunits in a pentamer. The carboxyl terminus of VP1 forms an extended arm that interacts with subunits of another pentamer, generating three kinds of interpentamer contact in the virus particle. This tying together of standard building blocks allows for the required variability in packing geometry without sacrificing specificity. Flexibly extended arms, which form ordered structures only when the units assemble into a particle, appear to be an important feature of complex assemblies.

A recently determined structure for the murine polyoma virus shows that an important surface loop is larger in the polyoma subunit than in SV40. Mutational evidence suggests that this loop creates a shallow pocket for binding sialic acid, required for cell entry by polyoma but not by SV40. A number of viruses of various structural types use cell-surface sialic acid for attachment, and it is of broad interest to understand how specific viruses accomplish the interaction. Comparison of SV40 and polyoma shows that this function can readily be added or lost by small changes at the surface of a viral coat protein.

We have recently begun to study the double-stranded RNA viruses. Crystals of rotavirus single-shelled particles and reovirus cores diffract to at least 7-Å resolution. These particles are elaborately organized molecular machines, containing the complete transcription and RNA-modification activities.

Receptors

The receptor for human immunodeficiency virus (HIV) is the lymphocyte surface antigen CD4. Its extracellular portion is composed of four immunoglobulin-like domains. We have determined the atomic structure of a two-domain, amino-terminal fragment, which binds HIV as tightly as does the intact receptor. The first two domains are joined by a continuous β-strand connector, and they have an extensive hydrophobic interface. Thus they form a rigid, rod-like segment. The HIV-binding site appears to be a ridge along one edge of the first domain. Binding and mutational studies, carried out collaboratively, show that a projecting phenylalanyl side chain is critical for the interaction with HIV gp120 and that various positively charged residues surrounding it are also important. The same region appears to be involved in contacts with class II MHC (major histocompatibility complex) molecules. We have also crystallized a four-domain fragment of CD4, corresponding to the entire extracellular region, and the structure determination is in progress.

Many important receptors are taken up into the cell by a process of endocytosis mediated by clathrin-coated vesicles. The transferrin receptor is one such molecule. It undergoes a well-characterized cycle of uptake and return to the cell surface. We have crystallized the extracellular domain, which makes up about three-fourths of the molecule. This domain exhibits reversible conformational changes at low pH that we believe to be significant for intracellular sorting steps. The determination of the structure has recently been facilitated by the observation that much better diffraction data can be collected by studying crystals at liquid-nitrogen temperatures. Such freezing techniques are now being broadly applied in our laboratory, enhancing our ability to study radiation-sensitive crystals of very complex structures.

Control of Gene Expression During the Cell Cycle and Development of the Mammalian Cerebellum

Nathaniel Heintz, Ph.D.—Investigator

Dr. Heintz is also Professor at the Rockefeller University. He received his Ph.D. degree at the State University of New York at Albany, where he studied the genetics and biochemistry of bacteriophage SPO1 gene expression. During postdoctoral studies with Robert Roeder at Washington University, St. Louis, he initiated his work on histone gene expression during the cell cycle. Continuation of these studies and examination of the developing mammalian cerebellum are his current research interests.

OUR studies are focused on the identification of molecular mechanisms controlling gene expression during the cell cycle and in the developing cerebellum. The elucidation of these mechanisms should provide fundamental insights into the biological transitions that underlie the control of cell division and the development of the mammalian central nervous system.

Control of Gene Expression During the Cell Cycle

Transcription of histone genes during the S phase of the eukaryotic cell cycle is achieved through the agency of subtype-specific consensus elements within histone gene promoters and their cognate distinct transcription factors. Coordinate transcription of this gene family is accomplished through biochemically distinct transcription factors; this suggests pleiotropic regulatory mechanisms that control the activities of these factors during the cell cycle. We wish to identify these mechanisms at the molecular level and elucidate their importance for cell-cycle progression. Identification of such mechanisms may provide highly specific targets for intervention in cell growth.

Recently we have examined the post-translational modifications to the transcription factor Oct1, which participates in histone H2b expression during the cell cycle. Both monoclonal and polyclonal antisera have been used to analyze changes in phosphorylation of Oct1 as cells proceed to division. We have discovered that multiple forms of Oct1 exist in the cell, and their distribution is dramatically regulated during the cell cycle.

Further analysis of Oct1 phosphorylation has established that late in the cell cycle this transcription factor appears to be a substrate for both CDC2 kinase and protein kinase A (PKA). Phosphorylation of Oct1 by PKA occurs within its DNA-binding domain and results in loss of DNA binding *in vitro*. This correlates with a significant decrease in Oct1 DNA binding during mitosis *in vivo*. These results indicate that modification of Oct1 during the cell cycle results in modulation of its function. Inactivation of Oct1 by phosphorylation during mitosis also provides a possible explanation of the long-standing observation that transcription is generally suppressed during this time in the cell cycle. Our present efforts are to extend this analysis to the histone H1 transcription factor H1TF2, to determine whether its control during the cell cycle may occur through the same molecular mechanisms. Demonstration that the timing and nature of the post-translational modifications on Oct1 and H1TF2 are similar *in vivo* would prove the existence of the proposed pleiotropic molecular mechanism for regulation of transcription during the S phase.

A question that has arisen from these studies is whether S-phase–specific transcription and DNA replication are mechanistically coupled. To address this issue we have focused on two specific questions: Are the regulatory proteins for S-phase histone gene transcription directly involved in DNA synthesis? Might proteins that regulate DNA synthesis at specific chromosomal origins of replication be activated by the same mechanisms that modulate those transcription factors?

In collaboration with Nicholas Heintz (University of Vermont) and Lisa Dailey (Rockefeller University) we have identified a cellular protein complex with several properties expected of replication-initiation factors. We have recently prepared antibodies and obtained primary amino acid sequences from one of these proteins (RIP60) and are using these tools to determine whether this factor participates in cellular DNA synthesis. Analysis of this factor during the cell cycle may help answer the two questions posed above.

Development of the Mammalian Cerebellum

The mammalian cerebellum is a complex and highly stereotyped structure in which major pattern formation and functional organization occur postnatally. The precise description of the cellular events occurring during cerebellar development, and the existence of many mutant mouse strains in which normal development of the cere-

bellum is perturbed, recommend it as an amenable system for molecular analysis of central nervous system development. Our initial interests in this area have been to identify genes that are either essential for normal development of the cerebellum or that serve as molecular markers for specific developmental events that occur during its formation. Our ultimate goal is to utilize these genes to identify novel proteins that are crucial to proper development of the cerebellum and to identify molecular mechanisms that participate in specific developmental events by analysis of the pathways that result in their correct spatial and temporal expression.

To identify genes that are required for normal development or maintenance of cerebellar structure and function, we have initiated efforts to clone the genes responsible for several neurological mutants of mice. Our most significant progress has been in studies concerning the *Lurcher* (*Lc*) and *meander tail* (*mea*) loci. *Lc* is a semidominant mutation that results in death of essentially all cerebellar Purkinje cells, beginning at about two weeks of age. Secondary loss of cerebellar granule cells and olivary neurons is also observed. We have constructed a detailed genetic map surrounding the *Lc* locus on chromosome 6 and have identified an RLFP (restriction fragment length polymorphism) marker approximately 0.5 cM from the gene. Genomic sequences from this closely linked marker were used to screen a yeast artificial chromosome (YAC) library from Shirley Tilghman (HHMI, Princeton University), resulting in isolation of a 280-kilobase YAC that maps to chromosome 6. Using sequences isolated from this YAC and informative recombinants generated during genetic mapping of the *Lc* locus, we have begun a chromosomal walk toward the *Lc* gene.

The gene *mea* is a recessive mutation resulting in gross perturbations of cerebellar cytoarchitecture that are confined to the anterior lobes of the cerebellum. The sharp boundary between the normal and affected area of the *mea/mea* cerebellum is reminiscent of the discrete boundaries evident in many *Drosophila* developmental mutants, suggesting that the *mea* gene may influence compartmental cellular organization in mammalian brain. In this case we have also constructed a detailed genetic map surrounding the *mea* gene on chromosome 4 and have begun efforts to identify appropriate genomic sequences to begin isolation of YAC clones containing the *mea* locus. The identification of genes responsible for these and other mouse neurological mutations should provide fundamentally important insights into cerebellar structure and function.

During the past year we have continued to pursue several different strategies to identify cDNA clones that are cell specific and developmentally regulated in the cerebellum. Using both subtractive hybridization and differential screening methods, a large number of novel cDNA clones have been isolated and are presently being analyzed.

One particularly successful strategy has focused on the use of antisera to purified granule neuron precursors to isolate genes that define stages in granule cell differentiation (in collaboration with Mary Beth Hatten, Columbia University). A large number of novel granule cell cDNAs have been isolated that define discrete stages of granule cell neurogenesis. One biological insight gained in these studies is the existence of a transient stage in development of granule cell neurons that occurs just as these cells complete their migratory journey from the external germinal layer to the internal granule layer of the developing mouse cerebellum. Since this step in granule cell differentiation has not been noted before, it will be interesting to determine its function, its relevance to neurogenesis in other areas of the central nervous system, and the role of the genes expressed uniquely at this stage in granule cell development.

Structural Biology of CD4 and CD8 Involvement in the Cellular Immune Response

Wayne A. Hendrickson, Ph.D.—Investigator

Dr. Hendrickson is also Professor of Biochemistry and Molecular Biophysics at Columbia University College of Physicians and Surgeons. He did his doctoral studies in biophysics at the Johns Hopkins University and remained for a year of postdoctoral research with Warner Love before going to the Naval Research Laboratory for continued postdoctoral study with Jerome Karle. He stayed on at NRL until he joined the faculty of Columbia University.

THYMUS-derived lymphocytes, or T cells, differentiate upon maturation into two major cell types. These are principally distinguished by the exclusive occurrence of either the CD4 or CD8 glycoproteins on their surfaces. The CD4-bearing cells are known as helper T cells, and the CD8-bearing cells as cytotoxic or killer T cells. T cells of both types are stimulated into action through the interplay of surface receptor molecules with peptide antigens, which must be presented on target cell surfaces as complexes with molecules of the major histocompatibility complex (MHC). The helper T cells ($CD4^+$ $CD8^-$) can only interact with class II MHC molecules, which occur on certain immune system cells such as macrophages and B cells, whereas the killer T cells ($CD4^-$ $CD8^+$) interact with the class I MHC molecules, found on all cells.

The involvement of CD4 or CD8 is essential for the efficient stimulation of the respective responses. Stimulation of helper T cells by peptides from an invading pathogen leads to the production of cytokines, which in turn stimulate the proliferation of selected lymphocytes for antibody production or for cellular defense. Subsequent stimulation of appropriate killer T cells by peptide antigens from infected cells then leads to the production of lytic factors that can eliminate the infected cell. Thus the molecules CD4 or CD8 can properly be described as co-receptors in the cellular immune response.

We have undertaken a series of structural studies on CD4 and CD8. This work is collaborative with Richard Axel (HHMI, Columbia University) and with Ray Sweet and his co-workers at SmithKline Beecham. Results from these studies, when taken in conjunction with mutational studies by others, are beginning to provide a picture of the molecular interactions implied by the specific stimulatory processes of the immune response. Moreover, the structural results also relate to the involvement of CD4 as a receptor for infection by the human immunodeficiency virus (HIV).

Crystals of Soluble CD4

Both CD4 and CD8 are single-pass transmembrane proteins. These can be made tractable for crystallographic study by the expression of soluble extracellular fragments of the protein. A soluble human CD4 fragment containing the entire extracellular portion of the protein was expressed in mammalian cells, and Peter Kwong has crystallized this protein into several different crystal lattices. Unfortunately, none of these diffracts very well. From a characterization of these crystals, it is possible, nevertheless, to deduce some interesting features of the CD4 molecule.

It appears that CD4 oligomerizes at the high concentrations needed for crystallization, and the lattice dimensions are consistent with a tetrameric model of 125 Å in length along a diad axis of the molecule. This propensity to associate may be relevant to signal transduction during T cell stimulation. Thus we are pursuing a crystallographic analysis of one of these crystals, despite the prospect of resolution limited to approximately 4 Å.

Structure of a CD4 Fragment

The consistently poor diffraction from several polymorphs of a chemically homogeneous protein preparation suggested to us that CD4 may be somewhat flexible. This possibility would be compatible with the four-domain structure deduced from sequence comparisons and intron positions, and limited proteolysis experiments produced stable fragments corresponding to the first two domains (D1D2) and the second two domains (D3D4). A recombinant construct expressing the D1D2 domain also produced a stable molecule, and Seong-Eon Ryu obtained diffraction-quality crystals of this fragment.

The resulting structure analysis revealed a D1 domain folded as in the variable domains of immunoglobulins and a D2 domain in a variation of the immunoglobulin constant-domain topology. These two domains are intimately connected, with the last strand of D1 running directly into the first strand of D2. Despite the similarity of D1 to immunoglobulin variable domains, there are appreciable differences in the loop structures.

One of these is the major determinant for HIV binding to CD4.

Structure of a CD8 Fragment

Although the function of CD8 is quite analogous to that of CD4, the architecture of the two molecules is different. CD8 is a disulfide-bridged dimer, whereas unactivated CD4 is monomeric, and each CD8 chain has only one immunoglobulin-like domain rather than four. This single domain is linked to the transmembrane segment by a 50-residue stretch that is highly glycosylated.

Dan Leahy has expressed a soluble fragment of CD8 composed of the immunoglobulin-like domain and half of the stalk region. He has recently solved a crystal structure of this molecule, and this shows that the CD8 dimer is organized very much as in the variable domain tips of antibodies. The stalk region from this molecule is disordered in these crystals, a feature compatible with a role as flexible tether to the membrane surface.

Interactions with MHC Molecules

Although a direct physical interaction between isolated CD4 or CD8 molecules and their respective MHC partners has not been demonstrated, these interactions are expected to be very weak. Otherwise, in light of the high polyvalency in cell-cell interactions, unwanted adhesion would be expected. Less-direct evidence for these interactions, however, is available from suitably constructed cell biology experiments.

In the case of CD4, Rafick Sekaly and his colleagues at the Clinical Research Institute of Montreal have used point mutations on human CD4 to identify residues involved in the interaction with class II molecules. These mutated residues map to exposed residues on one face of the D1D2 structural model. Further analyses are still in progress.

In the case of the CD8–class I interaction, we have been able to model a plausible mode of association between the positively charged CD8 molecule and the largely negative site localized to the $\alpha 3$ domain of the class I MHC molecule. This putative interaction is also consistent with results from a mutational study by Paula Kavathis at Yale University.

CD4 and AIDS

The hallmark of AIDS (acquired immune deficiency syndrome) is the immunodeficiency brought on by elimination of $CD4^+$ T cells. We continue in our efforts to understand molecular details of the interaction between CD4 and HIV that are involved in viral entry. Seong-Eon Ryu is completing his refinement of the D1D2 fragment, Reza Beigi is analyzing the structures of mutant proteins with abnormal HIV-binding properties, and Peter Kwong has characterized an antibody fragment that permits full HIV binding to CD4 but inhibits infection. This antibody, which binds to the D3 domain, indicates an involvement of CD4 flexibility in HIV entry. Peter Kwong is also attempting to crystallize the viral coat protein gp120 and various complexes with this crucial component of the system.

Variegated Position Effects in Drosophila

Steven Henikoff, Ph.D.—Investigator

Dr. Henikoff is also a member of the Basic Sciences Division of the Fred Hutchinson Cancer Research Center, Seattle. He received a B.S. degree in chemistry at the University of Chicago and a Ph.D. degree in biochemistry and molecular biology at Harvard University, working in the laboratory of Matthew Meselson. He did postdoctoral work with Charles Laird at the University of Washington.

EACH individual gene occupies a fixed position on a chromosome. By and large, moving a gene has only a minor effect on expression of the gene. Thus most studies of gene expression are able to focus on the gene as an independent unit, without taking into account larger organizational features. However, there are exceptional cases in which the relationship between a gene and its environment plays a role in expression of the gene.

The relationship between a gene and its chromosomal environment is especially apparent in examples of "position effects" associated with chromosomal rearrangements. In flies a well-known class of position effects involves inactivation of genes in the vicinity of rearrangement breakpoints. Gene inactivation is extremely variable from cell to cell, such that the affected tissue shows a variegated pattern of expression. In each case, it is found that the gene has been juxtaposed to heterochromatin, the deeply staining regions of chromosomes that flank the centromere. Although heterochromatin contains a substantial fraction of DNA in all higher eukaryotes, the repetitive sequence structure characteristic of heterochromatin and the near absence of genes have hampered attempts to understand its role in the genome. Genes that show variegated expression when placed next to heterochromatin provide a reporter function, allowing us to investigate these poorly understood regions of chromosomes.

Variegated position effects caused by juxtaposition to heterochromatin are seen for a large number of genes in *Drosophila*. One well-studied example is the *brown* gene, required for full pigmentation of the eye. Unlike nearly all other genes, however, such position effects on the *brown* gene are dominant over wild type—that is, placing one copy of *brown* next to heterochromatin can lead to inactivation of the other copy.

We have investigated the genetic basis for this gene inactivation in trans and have found that a necessary component is the pairing of homologues in the immediate vicinity of the *brown* gene. These findings have led to an explanation for "trans-inactivation," whereby protein components of heterochromatin make direct contact with the trans copy of the *brown* gene across paired homologues. In support of this hypothesis, we have been able to reproduce trans-inactivation at sites of transposons carrying the *brown* gene, but only for paired copies of the gene. In addition, we have found that even very small lesions that disrupt pairing in the immediate vicinity of the gene also reduce trans-inactivation.

Mapping of the sequence-specific component necessary for trans-inactivation to occur has localized it to the *brown* gene itself, probably to the region immediately upstream. This supports the notion that the contact between homologues is between a DNA-binding protein necessary for normal *brown* gene activity and a protein component of heterochromatin.

Our current efforts are aimed at identification of the protein components involved in trans-inactivation. One approach is to focus on the sequences in the immediate upstream region of the *brown* gene, where we expect that the sequence-specific component should bind. Another is to identify genes that encode proteins involved in the process by screening for mutations that specifically reduce the degree of trans-inactivation. We have now isolated several such mutations and are in the process of precisely mapping them to clone the corresponding genes.

A new research direction for us came from a serendipitous finding (during a screen for position-effect variegation mutations) of an unstable chromosome that causes gene markers carried on it to appear variegated. This chromosome derives from a fusion between a chromosome arm carrying the markers and a centromere from another chromosome. The unstable chromosome is intermediate in size among wild-type and rearranged linear *Drosophila* chromosomes, all of which are quite stable in somatic cells. The instability results from failure of the two products of replication—called sister chromatids—to come apart reliably at mitosis, leading to clones and single cells that have either gained an extra copy of the chromosome or have lost it entirely.

The appearance of the unstable chromosome

during mitosis indicates that instability is associated with premature separation of sister chromatids; as a result, the sisters might sometimes independently attach to the spindle apparatus that pulls them apart. Current evidence suggests that the defect results from a position effect on the centromere, because genetic suppressors of position-effect variegation also suppress the instability phenotype. To understand further the basis for instability, we have initiated physical mapping studies on this centromere. Such studies should help identify sequences necessary for centromere function in a higher eukaryote.

Phenomena that depend on the position of a sequence in the chromosome or on somatic pairing of homologues are easily observed in *Drosophila*, where powerful tools are available for genetic dissection. Related phenomena are known to occur in mammals, such as X chromosome inactivation, in which one of the female's X chromosomes becomes heterochromatic. The many similarities between chromosomes in organisms as diverse as flies and mammals lead to the expectation that an understanding of position effects and centromere function in *Drosophila* will have general implications.

Biological Roles and Expression of Complement Receptors

V. Michael Holers, M.D.—Assistant Investigator

Dr. Holers is also Assistant Professor of Medicine and Pathology at the Washington University School of Medicine and Assistant Physician at Barnes Hospital, St. Louis. He received his undergraduate degree from Purdue University and his M.D. degree from Washington University. He did postdoctoral research at the University of Colorado, Denver, and then at Washington University.

THE complement system was initially described as an activity found in serum that could mediate the killing of foreign infectious organisms such as bacteria or viruses. This system, in conjunction with specific immunity generated by vaccination, was found to be critical to preventing and fighting infections. Later it was realized that complement also facilitates the interaction of antigen-antibody complexes with cells of the immune system, which greatly enhances the specific immune response. Therefore complement not only helps to initiate an immune response but also plays an important role in the ability to clear infections and foreign antigens from the body.

The complement system consists of at least 20 serum proteins that are activated in a cascade fashion. As part of the activation process, protein fragments are released that attract inflammatory cells. In addition, antigen-antibody complexes are coated with specific complement fragments that covalently attach to this target. One of these fragments, complement component C3, is able to be cleaved proteolytically after attachment to a target. This cleavage reaction results in a number of different conformations of C3, which allow it to interact with at least three unique cell surface receptors. After binding to cells, specific transmembrane signals are sent, and the complement receptors then mediate ingestion and processing or killing of the targets. Also as part of this process, C3 fragments may bind to self tissues, rather than to the antibody-bound target, thereby attacking at inappropriate sites. Other cell membrane C3-binding proteins are able to inactivate this C3 and prevent inappropriate damage to self tissues.

We are interested in the interaction of C3 with its specific receptors and regulatory proteins, particularly the biological aspects of complement receptor 2 (CR2). In addition, we are studying mouse homologues of CR2 and other proteins of this type. Human CR2 serves as the receptor for the Epstein-Barr virus (EBV), which is responsible for most cases of infectious mononucleosis and is causally associated with a number of human tumors of B lymphocytes and epithelial cells. Patients who have forms of congenital or acquired immunodeficiency (such as AIDS or after organ transplantation) are particularly susceptible to tumors associated with EBV.

In the past few years we have cloned and analyzed the structure and activities of human CR2, its mouse homologues, and a unique mouse protein called Crry/p65. We have shown that expression of recombinant forms of these proteins in other cells is sufficient to mediate the binding of ligands or to control inappropriate complement deposition. By using other recombinant techniques and creating mutations within these proteins, we have shown that specific amino acids in small domains are important for ligand interactions. In addition, we have synthesized peptides that have the ability to block binding of some ligands, in particular, EBV binding to CR2. These studies should allow us to devise strategies to alter the function of this receptor *in vivo*. For instance, one type of reagent might block EBV binding to CR2 but not normal binding of C3. This could be useful in some illnesses associated with EBV.

We have analyzed the murine homologues of these proteins to understand further the biological role of human CR2, in addition to other complement receptors and regulatory proteins. Once the activities of these proteins are understood, we should be able to utilize murine models of the normal immune response, as well as autoimmune diseases, to elucidate the *in vivo* roles of these proteins.

Another aspect of CR2 expression is also under analysis. Expression of CR2 varies during human B lymphocyte development: it is expressed only on late pre–B cells and mature B cells and not on very early pre–B lymphocytes or on late immunoglobulin-secreting cells. The molecular mechanisms that underlie this phenotype, which is also found among other B cell–specific markers, are likely fundamental to the overall processes by which B cells mature and are activated. We are analyzing these mechanisms. As part of these studies we have defined a promoter for CR2 and other sites within the gene that are likely to be impor-

tant in gene regulation. We have determined that the receptor levels go up and down because the gene is turned on and off during development, we have identified functionally important domains and sites of protein interaction within the CR2 promoter, and we are clarifying the role of each site in CR2 expression and B cell maturation. These studies will increase our understanding of specific gene expression in B lymphocytes and further our knowledge of how to alter B cell phenotypes along pathways that might be more beneficial during certain disease states.

Genetic Control of Nematode Development

H. Robert Horvitz, Ph.D.—Investigator

Dr. Horvitz is also Professor of Biology at the Massachusetts Institute of Technology and Neurobiologist and Geneticist at Massachusetts General Hospital, Boston. He earned his undergraduate degrees in mathematics and in economics at the Massachusetts Institute of Technology, followed by the M.A. and Ph.D. degrees in biology from Harvard University. His postdoctoral research was done at the Medical Research Laboratory of Molecular Biology, Cambridge, England. Dr. Horvitz is a member of the National Academy of Sciences.

HOW do genes control animal development? Taking a primarily genetic approach to answer this question, members of our laboratory have isolated developmental mutants of the roundworm *Caenorhabditis elegans* and have used both genetic and molecular genetic techniques to characterize these mutants. Because the complete cellular anatomy (including the complete wiring diagram of the nervous system) and the complete cell lineage of *C. elegans* are known, mutant animals can be studied at the level of single cells and even single synapses. Genes that play specific roles in cell lineage, cell signaling, cell death, and cell migration have been identified and analyzed.

Cell Lineage

The problem of cell lineage—how a single fertilized egg cell undergoes a complex pattern of cell divisions to generate a multiplicity of distinct cell types—is one major focus of the research of our laboratory. We have identified hundreds of genes responsible for controlling aspects of the *C. elegans* cell lineage. Many of these genes function in generating cell diversity during development. For example, some genes act to make the two daughter cells generated by a single cell division different from each other, and one gene acts to make certain daughter cells different from their mothers. The action of some cell lineage genes is constrained to a single cell type, tissue, or organ. For example, one gene acts only in the nervous system, and another acts only in the hypodermis. Other genes act in multiple tissues.

We have analyzed a number of these genes at the molecular level. These studies have revealed that many genes that control cell lineage in *C. elegans* are strikingly similar to genes found in other organisms, including humans. Thus the analysis of developmental control genes in *C. elegans* should help us to understand aspects of the development of more-complex organisms.

Cell Signaling

Much of the development of *C. elegans*, like that of other organisms, involves intercellular communication. We have studied cell interactions in nematode development by using a laser microbeam to kill single cells in living animals: if destruction of one cell alters the fate of a second cell, the first cell must normally interact with the second. We have analyzed in detail the cell interactions involved in inducing the development of the vulva, which forms the external genitalia, connects the uterus with the outside environment, and is used for egg laying and copulation. We have characterized many genes that function in the cell interactions of vulval development. One gene that acts as a switch in the vulval inductive signaling pathway is a member of the *ras* gene family. Other *ras* genes are associated with many human cancers; the same mutations that cause extra vulval cell divisions in *C. elegans* are oncogenic in mammals. Another gene in the vulval signaling pathway has similarities to the *src* gene, which is also associated with human cancers. The study of these and other genes that function in cell signaling in *C. elegans* might provide insights relevant to cancerous growth in humans.

Cell Death

Naturally occurring or "programmed" cell death is common during the development of the nervous system of many animals, including *C. elegans*. Why organisms generate cells only to have them die is an intriguing question. Furthermore, the mechanisms responsible for cell death might be of medical importance, as the clinical features of many human disorders (including trauma, stroke, and a variety of neurodegenerative diseases) are a consequence of nerve cell deaths.

We have been identifying and characterizing genes that function in programmed cell death in *C. elegans*. Two genes cause cells to die, and seven other genes are involved in removing the corpses of dead cells. The two genes that cause cells to die must be expressed by the dying cells themselves, indicating that, at least to this extent, programmed cell deaths are cell suicides. Al-

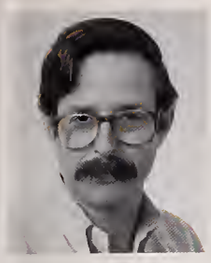
though many cells die during the course of *C. elegans* development, most cells survive; cell survival requires the inactivation of the cell death process, as a tenth gene functions to prevent the action of the nine cell death genes in surviving cells. In addition, this regulatory gene is itself controlled in a cell-specific fashion by other genes that determine which cells are to live and which are to die. Molecular analyses of genes involved in programmed cell deaths in *C. elegans* have suggested that the biochemical processes responsible might be similar to those suspected to cause nerve cells to die in human neurological disorders. We hope that knowledge of what makes cells die and of what can block the cell death process in *C. elegans* will lead to methods that will prevent the cell deaths responsible for human disorders.

Cell Migration

During animal development, cells are often generated far from their final positions and must migrate considerable distances before being able to function. To understand what causes cells both to migrate and to stop migrating, we are analyzing two *C. elegans* cell migrations. The first involves a pair of muscle precursor cells that are born in the posterior body region and move to a central position along the animal's length, near its gonad. We have discovered that these migrations involve signaling between the migrating cells and gonadal cells located at the termination site of the migration. We are characterizing genes that function in this signaling process.

The second migration we have studied involves a pair of neuronal cells that move from the tail region to the midbody region of the animal. Thirteen genes have been identified that must function for these neuronal migrations to occur properly. Some of these genes probably act in the migration process *per se*, but some do not. The actions of some of these latter genes allow these neurons to acquire their identities; if these genes do not function, these neurons fail to express their normal characteristics, including their long-range cell migration.

Protein Folding in Vivo



Arthur L. Horwich, M.D.—Associate Investigator

Dr. Horwich is also Associate Professor of Human Genetics and Pediatrics at Yale University School of Medicine. He received A.B. and M.D. degrees in biomedical sciences from Brown University. His internship and residency training in pediatrics were done at Yale. His postdoctoral research training was at the Salk Institute with Walter Eckhart and at Yale University with Leon Rosenberg.

UNTIL recently it has been assumed that newly made proteins in the living cell, comprising amino acid chains with characteristic sequences, are able to fold spontaneously into precise three-dimensional structures that exhibit biological activity. Such folding has been observed in test-tube experiments, where many proteins unfolded by denaturing agents can be diluted from these agents and observed to refold into their biologically active forms. Recent studies, however, suggest that in the cell the process of folding is assisted by other specialized proteins. We originally identified such a protein while studying mitochondria, the intracellular organelles that carry out energy metabolism.

Most of the proteins of mitochondria are first made outside the organelles, in the cytosol, and then imported through two membranes to reach the innermost mitochondrial "matrix" compartment. To traverse the membranes, the newly made proteins are first unfolded on the cytosolic side. After import, the proteins refold on the inside of the organelles into their biologically active conformations. We identified a mutant cell in which mitochondrial proteins were imported but failed to fold into biologically active forms. The mutation was found to affect a protein that normally resides in the matrix compartment, called heat-shock protein 60 (hsp60).

This protein was originally identified by the observation that its abundance is increased about twofold in response to incubation of cells at high temperatures. However, it is an abundant protein even before heat shock, and our genetic analysis demonstrated that, consistent with a critical baseline function, hsp60 is required not only at high temperatures but at all temperatures. The increased level produced in response to heat stress could represent an effort to try to refold efficiently mitochondrial proteins that heat has denatured.

In the mitochondrial matrix, hsp60 is found in a higher order structure, a complex. Fourteen copies of the protein are arranged in two stacked rings, a "double donut." Each ring contains seven radially arranged copies of hsp60. Our studies have demonstrated that unfolded mitochondrial proteins entering the matrix space become associated with the surface of the hsp60 complex. Then, in steps requiring both energy (ATP) and a second protein, the polypeptides are folded into their active forms and released from the complex.

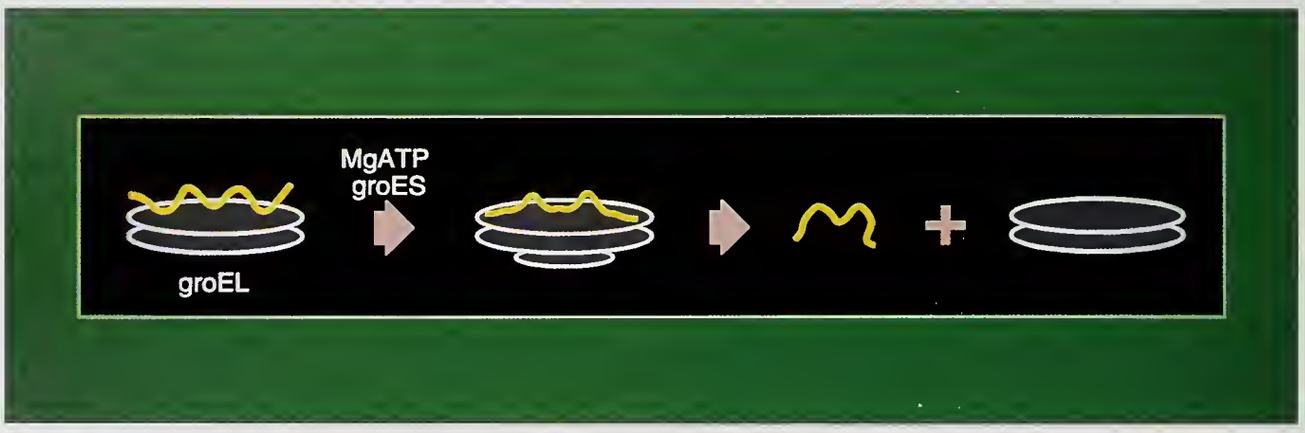
The folding pathway must be dictated by the amino acid sequence of the "substrate" protein to be folded, not by the hsp60 complex, because we have used the complex to fold proteins that normally reside outside mitochondria. It seems, however, that hsp60 acts by speeding up, or catalyzing, the folding of proteins. How does it do this? One possibility is that it simply prevents domains of proteins from wrongfully interacting, either with each other or with nearby proteins in the mitochondrial matrix, a "chaperone" function. Another possibility is that the complex actively promotes the progression of an unfolded protein through a series of folding steps.

Because mitochondria arose from bacteria (one cell ingested another), it is not surprising that a structurally related component, the groEL protein, has been found in bacteria. In the 1970s it was observed that when *Escherichia coli* cells partially defective in this protein are infected with a virus, the newly made viral coat proteins are unable to assemble to make new virus particles.

Additional evidence comes from our reconstitution experiments. When a test protein unfolded in denaturant was diluted into a mixture containing purified groEL complex, it became bound to the complex and was thus prevented from misfolding and aggregating. The bound protein was found as a folding intermediate, called a molten globule, that has formed its local structures but lacks the three-dimensional organization of the active form. When both a cooperating component, groES (a small ring structure also with seven members), and ATP were added, the polypeptide was observed to reorganize its structure while in association with the complex during the next minutes and to be released in its active form. The cost of folding a single polypeptide was approximately 100 ATP hydrolyzed, amounting to about

one-tenth of the amount of energy needed to synthesize the polypeptide.

How general is the utilization of folding machinery outside bacteria and mitochondria? We thought we might find a functionally similar component in the cytosol of higher organisms by looking first in the kingdom of archaebacteria. These organisms are evolutionarily distinct from bacteria and have recently been shown to contain components closely related to those in the cytosol of higher organisms. We transferred a thermophilic archaebacterium that grows normally at 75°C to a near-lethal temperature of 88°C and found that a single major heat-shock protein was produced. This protein, already abundant at normal temperature, is a double-ring complex that looks much like the complexes of mitochondria and bacteria, except that each ring contains nine members. The purified ring complex binds unfolded proteins and exhibits ATPase activity, consistent with the idea that it might function in the archaebacteria similarly to the hsp60 and groEL complexes. Analysis of the protein's amino acid sequence demonstrated no significant relationship to either hsp60 or groEL, but rather a relationship to a protein called TCP1, an essential cytosol protein of higher organisms, whose function has been largely unknown but which has been implicated in assembly of the mitotic spindle apparatus. Using a mutant yeast cell defective in TCP1, we are now investigating whether the protein has a specialized function in assembly of the spindle or a more general function as a "folding machine" for the cytosol of higher organisms.



Model for folding of polypeptides by groEL. Polypeptide, shown in yellow, is bound by groEL and stabilized in "molten globule" conformation. When MgATP and groES are present, polypeptide undergoes conformational changes and is ultimately released, reaching a native, active conformation.

Derived from Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A.L., and Hartl, F.-U. 1991. Nature *352:36–42.*

Molecular Mechanisms in the Regulation of Synaptic Transmission

Richard L. Huganir, Ph.D.—Associate Investigator

Dr. Huganir is also Associate Professor of Neuroscience at the Johns Hopkins University School of Medicine. He completed his undergraduate work in biochemistry at Vassar College and received his Ph.D. degree in biochemistry and molecular and cell biology from Cornell University, where he performed his thesis research in the laboratory of Efraim Racker. After completing a postdoctoral fellowship with Paul Greengard at Yale University School of Medicine, Dr. Huganir moved to the Rockefeller University.

INFORMATION processing in the brain depends on the transmission of signals between neurons at specialized areas of contact, called synapses. At synapses, ion channel proteins in the neuronal cell membrane generate an electrical current, which triggers the release of chemical signals from the first neuron, called the presynaptic neuron. These chemical signals, or neurotransmitters, bind to specific receptor proteins in the membrane of the second neuron, called the postsynaptic neuron. The neurotransmitter receptors then generate electrical currents in the postsynaptic neuron and thus complete the process of synaptic transmission.

The efficiency of synaptic transmission at any given synapse is constantly changing in response to a variety of factors; this synaptic plasticity plays a major role in the function of the nervous system. Both the amount of neurotransmitter released from the presynaptic neuron in response to a given electrical signal and the sensitivity of the postsynaptic receptor system for a given amount of neurotransmitter can be modulated. The molecular mechanisms that underlie the modulation of synaptic transmission have only begun to be defined. Recent studies have provided evidence that protein phosphorylation is an important mechanism in the regulation of synaptic transmission.

Protein phosphorylation systems consist of three primary components, a protein kinase, a substrate protein, and a phosphoprotein phosphatase. Protein kinases are enzymes that catalyze the chemical transfer of phosphate molecules from ATP to specific substrate proteins. The activities of many protein kinases are regulated by neurotransmitters and hormones through the actions of substances called second messengers, such as cAMP, calcium, and diacylglycerol. Substrate proteins include many cellular components, among them enzymes, ion channels, and neurotransmitter receptors. The addition of the negatively charged phosphate group alters the structure of these substrate proteins, thereby regulating their functional properties. Phosphoprotein phosphatases are enzymes that reverse the process of protein phosphorylation, remove the phosphate group from the substrate protein, and return it to its basal state.

My laboratory is concerned with the structure and function of neurotransmitter receptors and the role of protein phosphorylation in the regulation of the properties of the neurotransmitter receptors. We have used as a model system the best-characterized neurotransmitter receptor and ion channel in neurobiology today, the nicotinic acetylcholine receptor. In addition, we have been studying the major inhibitory neurotransmitter receptors in the brain, the $GABA_A$ receptors, and the major excitatory neurotransmitter receptors in the brain, the glutamate receptors. These receptors are neurotransmitter-dependent ion channels that generate electrical currents in the postsynaptic membrane of the synapse in response to their neurotransmitter.

To study the molecular mechanisms involved in neurotransmitter receptor and ion channel function, it is essential to identify chemically the specific proteins required for this activity. We began by defining the molecular components required for the functioning of the nicotinic acetylcholine receptor ion channel. Using membrane reconstitution techniques, we solubilized the nicotinic receptor and its ion channel from isolated postsynaptic membranes, purified it, and reconstituted it into phospholipid vesicles. These studies demonstrated that the purified receptor, consisting of four types of protein subunits (α, β, γ, δ), contains the ion channel and has all the biological properties of the nicotinic receptor in the intact cell.

We next began to characterize the protein phosphorylation of these structural components. Postsynaptic membranes isolated from synapses highly enriched in the nicotinic acetylcholine receptor contain at least three different types of protein kinases that phosphorylate the nicotinic receptor on six different phosphorylation sites: cAMP-dependent protein kinase phosphorylates the γ- and δ-subunits of the receptor; a calcium- and diacylglycerol-dependent protein kinase phosphorylates the δ-subunit; and a protein-

tyrosine kinase phosphorylates the β-, γ-, and δ-subunits. These postsynaptic membranes also contain phosphoprotein phosphatase activity that dephosphorylates the phosphorylated nicotinic acetylcholine receptor.

We are currently using protein purification and molecular cloning techniques to characterize the protein-tyrosine kinases that phosphorylate the receptor and the phosphotyrosine protein phosphatases that dephosphorylate the receptor. We recently identified several cDNA clones for different types of protein-tyrosine kinases that are expressed in cells enriched in the nicotinic receptor and are attempting to determine which of these protein-tyrosine kinases phosphorylate the receptor. In addition, we recently purified the phosphotyrosine protein phosphatase that dephosphorylates the tyrosine-phosphorylated acetylcholine receptor and are using molecular cloning techniques to isolate cDNA clones for this phosphotyrosine protein phosphatase.

What are the functional effects of phosphorylation of the receptor by these protein kinases? We have examined this question directly by studying the properties of the purified and reconstituted receptor phosphorylated to different degrees by the various protein kinases. Phosphorylation of the receptor on the γ- and δ-subunits by cAMP-dependent protein kinase or phosphorylation of the receptor on the β-, γ-, and δ-subunits by the protein-tyrosine kinase dramatically increases the rate of desensitization of the receptor. Desensitization is the process by which the receptor is reversibly inactivated in the continued presence of the neurotransmitter acetylcholine. These studies provide direct evidence that protein phosphorylation of the nicotinic acetylcholine receptor regulates its physiological properties and plays a role in modulating its sensitivity to acetylcholine.

To analyze the effect of phosphorylation on the desensitization of the receptor in more detail, we have recently used site-specific mutagenesis techniques to mutate the phosphorylation sites on the receptor subunits. Mutant receptor subunits lacking phosphorylation sites have been expressed in *Xenopus* oocytes, and the regulation of desensitization of these receptors by protein phosphorylation is being analyzed and compared with normal receptors.

Using muscle cell cultures that are highly enriched in the acetylcholine receptor, we have investigated the regulation of the phosphorylation of the receptor by neurotransmitters, hormones, and neuropeptides. Calcitonin gene-related peptide (CGRP), a neuropeptide that is released from the presynaptic neuron with acetylcholine, increases the intracellular levels of cAMP and thereby regulates the phosphorylation of the receptor by the cAMP-dependent protein kinase. In addition, studies in our laboratory suggest that acetylcholine itself regulates intracellular levels of calcium and thereby regulates the phosphorylation of its own receptor by the calcium- and diacylglycerol-dependent protein kinase. We have also demonstrated that tyrosine phosphorylation of the nicotinic receptor is regulated by the neurons that synapse on muscle. More recently, in collaboration with Bruce Wallace (University of Colorado Health Sciences Center), we have found that agrin, an extracellular matrix protein, may be the factor from neurons that regulates tyrosine phosphorylation of the receptor. Agrin is a well-characterized protein that is secreted from neurons and induces receptor clustering under the nerve during synapse formation. These results suggest that agrin-induced tyrosine phosphorylation of the receptor may be involved in the induction of clustering of the receptor at the synapse.

Our recent studies on $GABA_A$ and glutamate receptors have supported our hypothesis that protein phosphorylation of neurotransmitter receptors plays an important role in the modulation of their function. We have expressed the genes for these receptors in a variety of cells and have analyzed the effect of phosphorylation on their functional properties. Using this system, we have shown that phosphorylation of the β-subunit by cAMP-dependent protein kinase decreases the response of the $GABA_A$ receptor to its neurotransmitter and alters the desensitization of the receptor. In addition, recent studies have demonstrated that glutamate receptors are phosphorylated by a protein-tyrosine kinase. The functional effect of tyrosine phosphorylation of the glutamate receptor is currently being examined. This work, combined with our studies of the nicotinic acetylcholine receptor, provides strong evidence that protein phosphorylation of neurotransmitter receptors is a primary mechanism in the regulation of synaptic transmission.

Molecular Aspects of Signal Transduction in the Visual System

James B. Hurley, Ph.D.—Associate Investigator

Dr. Hurley is also Associate Professor of Biochemistry at the University of Washington School of Medicine. He received his undergraduate degree in chemistry from the State University of New York College of Environmental Science and Forestry, Syracuse, and his Ph.D. degree in physiology and biophysics from the University of Illinois, Urbana, where he worked with Thomas Ebrey. His postdoctoral research included studies with Melvin Simon at both the University of California, San Diego, and the California Institute of Technology, and with Lubert Stryer at Stanford University.

OUR laboratory studies molecular mechanisms responsible for visual transduction in vertebrate and invertebrate photoreceptors. Despite the fact that these two types of cells respond to light via quite diverse mechanisms, they have many general features in common. We are investigating mechanisms that determine such photoreceptor characteristics as sensitivity, rates of activation and deactivation, and ability to adapt to constant light.

Light hyperpolarizes vertebrate photoreceptors via a G protein–mediated cascade that culminates in cyclic GMP hydrolysis. Depletion of cGMP reduces the activity of cGMP-gated cation channels in the photoreceptor plasma membrane. In darkness Ca^{2+} enters the cell through these channels. Light blocks this entry, and the resulting depletion of cytosolic Ca^{2+} promotes recovery by stimulating guanylate cyclase to resynthesize cGMP.

Invertebrate photoreceptors respond to light very differently. In these cells light activates phospholipase C, which produces inositol triphosphate and diacylglycerol as second messengers. Few biochemical details of invertebrate phototransduction are well understood.

Vertebrate Phototransduction

Our laboratory recently identified a novel Ca^{2+}-binding protein that imparts Ca^{2+} sensitivity to photoreceptor guanylate cyclase. This protein, named recoverin, promotes recovery by stimulating guanylate cyclase when free Ca^{2+} concentrations fall below 300 nM. The amino acid sequence of recoverin reveals three Ca^{2+}-binding sites. Ca^{2+} influences a variety of physical properties of recoverin, including fluorescence and mobility on electrophoresis gels. We cloned recoverin cDNA and expressed recombinant recoverin in *Escherichia coli*.

The effects of Ca^{2+} on recombinant recoverin and retinal recoverin are quite different. To account for these differences, we compared the masses of recombinant and retinal recoverin directly by ion-spray mass spectrometry. To our surprise, the modification turned out to be a novel type of heterogeneous amino-terminal acylation. Each recoverin is acylated with either a C14:0, C14:1, C14:2, or C12:0 fatty acid residue.

Following stimulation by light, transducin, the photoreceptor G protein, hydrolyzes its bound GTP and loses its ability to activate phosphodiesterase. Photoreceptor cells recover from a light flash within a couple of seconds, but the steady-state hydrolysis of GTP is slower. To clarify the role that GTP hydrolysis plays in recovery from a photoresponse, we produced transgenic mice that express a mutant transducin that hydrolyzes GTP more slowly than their normal counterparts. Preliminary results suggest that photoreceptors expressing this form of transducin are abnormally desensitized. This effect may reflect an attempt by the cells to compensate for the persistent phosphodiesterase activity of the mutant transducin.

Drosophila Vision

Biochemical and physiological evidence suggests that a G protein mediates invertebrate phototransduction by stimulating phospholipase C. We characterized several *Drosophila* G proteins with the aim of understanding their role in invertebrate phototransduction. In addition to several G protein α-subunits, two G protein β-subunits were identified. The first one was detected throughout the nervous system but not in the eyes. This prompted us to search for a photoreceptor G protein β-subunit.

Through use of a specific type of cDNA screening method, we were able to identify a novel type of G protein β-subunit that is expressed specifically in the *Drosophila* compound eye. Recently, in collaboration with Charles Zuker and his colleagues (HHMI, University of California, San Diego), two mutant *Drosophila* strains have been identified with reduced expression of this eye-specific β-subunit. Biochemical analyses of eyes from normal *Drosophila* and from these mutants are being used to study the role of G proteins in phototransduction.

A toxin from the microorganism *Bordetella pertussis* specifically inactivates certain G proteins. To study the physiological importance of G proteins in *Drosophila*, we produced transgenic flies that conditionally express endogenous pertussis toxin. Pertussis toxin induction in adult *Drosophila* alters their visual response and eating behavior. We are investigating the molecular mechanisms responsible for these phenotypes.

The long-term objective addressed in these projects is to identify biochemical mechanisms by which photoreceptors respond to stimuli.

The Molecular Basis of Cell Adhesion in Normal and Pathological Situations

Richard O. Hynes, Ph.D.—Investigator

Dr. Hynes is also Professor of Biology and Director of the Center for Cancer Research at the Massachusetts Institute of Technology. He received his undergraduate degree in biochemistry from the University of Cambridge and his Ph.D. degree in biology from the Massachusetts Institute of Technology. After several years of postdoctoral work at the Imperial Cancer Research Fund Laboratories in London, where he initiated his early work on fibronectins, he returned to MIT as a faculty member. Dr. Hynes has been the recipient of a Guggenheim Fellowship and is a Fellow of the Royal Society of London and the American Association for the Advancement of Science.

MOST cells in the body adhere to their neighbors and to the extracellular matrix, a complex array of proteins that comprise a fibrillar meshwork throughout the body. Cell adhesion plays important roles in the normal functions of cells, contributing to cellular organization and structure, proliferation, and metabolism. During embryological development, cell adhesion is important for the movements of cells that contribute to modeling of the embryo. In the adult, appropriate cell adhesion is necessary for numerous physiological processes.

For example, in the blood, cells known as platelets adhere to the walls of blood vessels that are damaged and help to prevent bleeding. This adhesion process is essential to protect against hemorrhage. On the other hand, it is equally important that platelets should not adhere at inappropriate times. If they do, the result is thrombosis. Thus the control of platelet adhesion is a matter of life and death. Other blood cells involved in defense mechanisms during infection or inflammation need to adhere to the walls of blood vessels at the sites of infection to emigrate into the affected tissues.

Another process involving cell adhesion and migration is wound healing. When skin is damaged, the skin cells migrate in over the wound to cover it. The processes of cell migration involved in wound healing have much in common with those occurring during development.

A final example is that of cancer. Tumor cells exhibit altered adhesion both to one another and to their surroundings. This altered adhesion is thought to be involved in the invasion and metastasis of tumor cells.

These examples illustrate the importance of appropriate adhesion of cells to their surroundings. Our laboratory is involved in molecular analyses of these processes. We seek to understand the proteins involved in cell adhesion and how they control adhesion and migration of cells in both normal and pathological processes.

Two main classes of proteins interest us. The first comprises the large proteins that make up the extracellular matrix. These proteins cooperate to build a fibrillar meshwork to which the cells attach and on and through which they migrate. We have investigated several of these proteins, which we refer to as "nectins" to denote their role in binding to cells. Fibronectins, a closely related group of proteins all encoded by a single gene, are the best understood of these nectins. The different forms of fibronectin are generated by alternative RNA splicing. We and others have analyzed in detail the functions and the structure of these proteins. This work is leading to a deeper understanding of their roles in cell behavior. For example, it is now known that fibronectins have several sites in each molecule that bind cells. The detailed structure of these binding sites is being elucidated. One intriguing observation is that fibronectins share with many other nectins a common recognition site made up of only three amino acids. This site (designated RGD in the single-letter amino acid code) is recognized by receptor molecules on cell surfaces. This interaction can be blocked by antibodies to the nectins or to the receptors, which are known as integrins, or by competitor peptides containing the RGD sequence. Such blockades interfere with the cell-adhesive interactions involved in the physiological processes discussed above. Recent work has identified other cell-binding sites within fibronectins, which are recognized by different receptors. Cells interact with these molecules in a complex fashion, which is as expected, given the participation of cell adhesion in many diverse cellular functions.

Our second major focus of interest is the family of integrin receptors. These comprise a family of related cell surface receptors, each composed of two subunits. Each integrin receptor has a particular specificity for certain nectins and mediates the interactions of cells with the extracellular matrix. In addition, the integrins connect to the inside of the cell, where they mediate interactions with the internal structures or cytoskeleton of the cell that are involved in the shape, organization, and migration of cells. This integration of the organization of the extracellular matrix with

the cytoskeleton inside the cells is one of the origins of the name "integrins."

Using the methods of cell and molecular biology, we are studying the structure and function of fibronectins and integrins, their interactions, and their roles in various physiological processes, including development, blood clotting, inflammation, wound healing, and cancer. We observe regulated expression of these molecules during these processes, and it is clear that these molecules are crucial for the appropriate behavior of cells. For instance, altered expression of both fibronectins and integrins in tumor cells contributes to their wayward behavior, and expression of these proteins is altered during wound healing. Using recombinant DNA methods, we can produce specific and modified forms of fibronectins and integrins and thus investigate the ways in which they affect the behavior of individual cell types.

We have recently made progress in analyses of the role of the intracellular portions of various integrin receptors in interactions with the cytoskeleton. We have also obtained detailed structural information about the cytoskeletal protein talin, which is a primary candidate for interactions with integrins. We are now investigating possible interactions between normal and mutant integrins and talin. Further progress along these lines should help explain the effects of cell adhesion on cell structure and behavior.

Work in the past year has uncovered evidence that integrins do more than provide a physical connection between the extracellular matrix and the cytoskeleton. Engagement of integrins triggers tyrosine phosphorylation inside the cells, strongly suggesting signaling via integrins. Integrin function can also be regulated from inside cells. Cells need to detach as well as attach. How is this regulated? We find that cell detachment can apparently be triggered by phosphorylation of the cytoplasmic domains of integrins. Thus our current picture of integrins is that they can mediate signaling both into and out of cells.

To extend our understanding of the roles of fibronectins and integrins to intact organisms, we use genetic analyses in two animal systems. First, we are analyzing the role of integrins during the development of *Drosophila melanogaster*, a fruit fly that is suitable for genetic analyses. Flies with mutations in genes encoding integrins have defects in embryonic development, in muscle function, and in the development of wings and eyes. Analyses of these defects provide insights into the functions of these proteins. We are also identifying new integrin species in *Drosophila*.

In the second genetic project, we have generated strains of mice that are mutant for fibronectins. When both normal copies of the fibronectin gene are ablated, mouse embryos cannot proceed normally beyond the early developmental stage known as gastrulation. We have also made more-subtle mutations in the fibronectin gene and are currently analyzing their effects. These mutant mice, together with transgenic mice expressing different forms of fibronectin, should allow us to dissect the functions of the various forms of fibronectin *in vivo*. Encouraged by the progress on fibronectin molecular genetics in mice, we have also begun to generate mutations in integrin receptor genes and in other cell adhesion molecules, particularly selectins, cell-cell adhesion molecules involved in the early steps of inflammation. As our analyses proceed, we will be able to investigate the effects of the various mutations on hemostasis, thrombosis, wound healing, and tumor development.

These studies should provide a deeper understanding of the molecular basis of cell adhesion and its involvement in physiological and pathological processes. This understanding, in turn, should provide opportunities for therapeutic treatments of diseases such as thrombosis and cancer.

The work on cultured cells and some of the work on mice are supported by grants from the National Institutes of Health.

Molecular Genetics of Intracellular Microorganisms

Ralph R. Isberg, Ph.D.—Assistant Investigator

Dr. Isberg is also Assistant Professor of Molecular Biology and Microbiology at Tufts University School of Medicine. He received his A.B. degree in chemistry from Oberlin College and his Ph.D. degree in microbiology and molecular genetics from Harvard Medical School. He conducted postdoctoral work on bacterial pathogenesis in the laboratory of Stanley Falkow at Stanford University. His honors include a Searle Scholars Award and a National Science Foundation Presidential Young Investigator Award.

MANY species of bacteria are capable of causing diseases by colonizing and growing within human hosts, using tactics that avoid normal immune responses. As part of a general strategy to establish an infectious niche, a variety of microorganisms cause diseases by entering and growing inside human cells soon after encounter. Bacteria that establish infections in this manner are called intracellular microorganisms. Among the diseases they cause are tuberculosis and the most common types of sexually transmitted and food-borne diseases found in the industrialized world. Despite the prevalence of such infections, there was little information until recent years on the factors expressed by these microorganisms that allow them to enter host cells and thrive.

The objectives of our research are to investigate two important aspects of the life-style of intracellular microorganisms. First, we would like to determine at the molecular level how these organisms can enter human cells that do not normally internalize bacteria. Second, we want to analyze factors they encode that allow them to survive and grow within the ordinarily hostile environment of human cells. Our main approach has been to identify bacterial species that enter or grow particularly well within host cells and to develop genetic and biochemical techniques for analyzing their strategies. The primary rationale for this approach is that it provides insights into basic processes that are applicable to numerous intracellular microorganisms.

To investigate the molecular mechanism of bacterial binding and entry into host cells, we have been analyzing the bacterium *Yersinia pseudotuberculosis*, an organism that causes an intestinal disease often accompanied by infection of multiple organ systems.

To investigate intracellular growth, we have been analyzing *Legionella pneumophila*, the causative agent of Legionnaire's disease pneumonia. The intracellular growth process of the bacterium is very similar to that of a wide range of intracellular microorganisms, and development of molecular strategies for analyzing it has been relatively straightforward.

Yersinia pseudotuberculosis Entry Into Cultured Human Cells

Y. pseudotuberculosis can enter host cells via three different paths. For each path the microorganism apparently encodes a unique set of protein factors to be used at different tissue sites during the infection process. We have focused on the path that is promoted by the protein invasin, the product of the bacterial *inv* gene. Invasin is a 108-kDa protein on the surface of the bacterium that allows it to enter human cells by binding receptor molecules on their surface. We have shown that a 20-kDa region of invasin binds the host cells, and this region is sufficient to promote uptake. Evidence indicates that after the binding occurs, the host cells do most of the work in internalizing the bacterium.

Invasin binds at least four different receptors. Called integrins, these had been previously identified by investigators interested in a variety of mammalian cell adhesion processes. The particular integrin receptors that bind invasin can adhere to a variety of mammalian proteins, such as fibronectin and molecules that allow adhesion of immune response cells to inflamed tissues.

Although invasin binds these well-characterized receptors, there is no obvious sequence similarity between invasin and other proteins that bind integrins, and mutations that eliminate the interaction between invasin and its receptors identify amino acid residues not previously shown to be involved in integrin binding.

Integrins are clearly not the only host-encoded factors necessary for internalization. Analysis of this process has indicated that two additional factors are required. Mutant studies of one of these integrin receptors indicate that a cell structure called a clathrin coat directly interacts with the integrin receptor during the internalization of the bacterium. If this interaction is eliminated, the bacterium cannot be internalized. A second structure, the host cell cytoskeleton, also is involved in the internalization, but we believe that this structure performs an indirect role and does not directly bind the integrin during all stages of bacterial uptake.

Our investigation of invasin/integrin has led to a model for *Yersinia* uptake into host cells. Binding of invasin to its integrin receptor leads to rearrangement of the cytoskeleton—rearrangement requisite to entry. A signal must be sent to cause the host cell to internalize the microorganism, and the internalization is facilitated by the extraordinary avidity with which invasin binds its receptors. Other proteins that bind the identical integrins cannot produce this signal so efficiently, because they do not bind the receptors tightly.

Thus invasin appears to promote entry of the microorganism because it binds an important receptor that interacts with clathrin coats and communicates with the cell cytoskeleton, and because it binds so tightly to this receptor.

Legionella pneumophila Growth in Phagocytic Cells

L. pneumophila causes a variety of diseases in humans, including Legionnaire's disease pneumonia. The bacterium grows in lung tissues after encounter with its human host. Its favorite habitat is within alveolar macrophages, cells that normally function to kill invading microorganisms. An important mechanism for macrophages to kill or inhibit the growth of a microorganism is to internalize it and sequester it in a compartment called a phagosome, which in turn fuses with a lysosomal compartment filled with antibacterial factors. *L. pneumophila* is able to grow within the phagosome, convert it into an organelle with a unique morphology, and prevent the introduction of the antibacterial lysosomal components into this site.

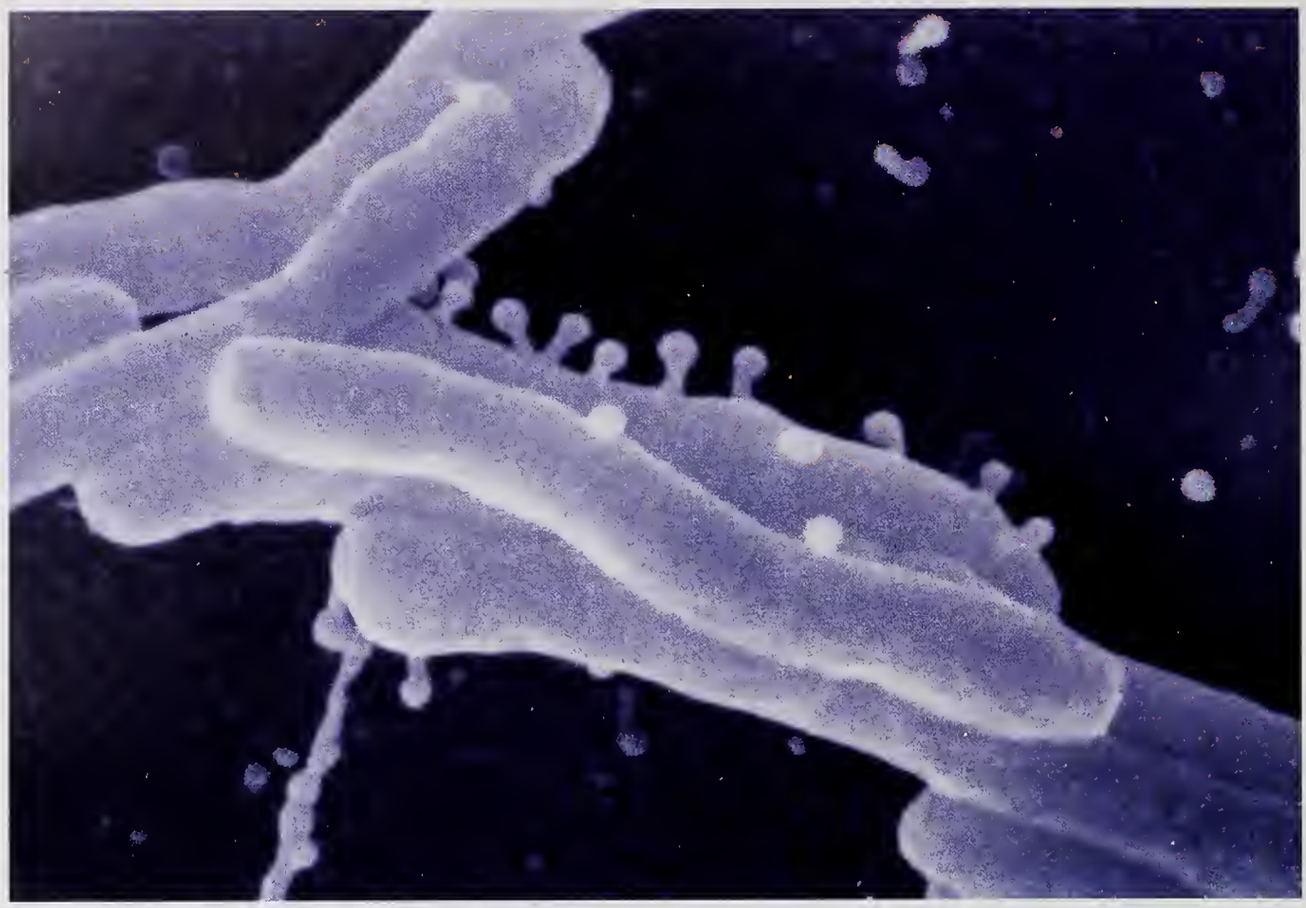
We have been interested in determining how *L. pneumophila* is able to establish and grow within this protective niche. We have been taking two tactics toward analyzing this process. Our first approach has been to isolate mutations in this bacterium that prevent it from growing intracellularly. Our second approach has been to identify factors that are selectively synthesized by the bacterium only during intracellular growth.

Using the first approach, three easily distinguishable classes of mutants have been isolated. The first class causes the bacterium to be internalized by a macrophage via a novel pathway, and this causes an extreme defect in bacterial growth. The second class, and most easily isolated, consists of mutants that are no longer able to prevent the lysosomal contents from being introduced into the phagosome. The third class appears normal for uptake as well as for shutting out the lysosomal components, but the phagosome containing the mutant microorganism no longer exhibits the unique morphology usually found in a *Legionella* infection.

These classes of mutants indicate that the microorganism performs a distinct series of steps within the macrophage, each of which contributes to the parasite's efficient growth. To investigate the factors missing in these mutants and analyze the steps in growth performed by this bacterium, we have identified a small region of the *Legionella* chromosome that encodes the factors missing in the latter two mutant classes. We are currently analyzing this region of the chromosome intensively, with the hope of purifying the factors encoded by this region in order to describe their functions in molecular detail.

Our second approach involves using a novel scheme to identify *L. pneumophila* genes that are regulated in a fashion such that they are turned off when the bacterium is growing outside the host cell but are rapidly turned on during growth within macrophages. This tactic involves cloning fragments of *L. pneumophila* chromosomal DNA in front of a reporter gene and introducing these molecular clones into the bacterium. All molecular clones constructed in this manner are killed unless they contain genes that are regulated in the desired fashion. Using this strategy, we have cloned at least eight genes that are normally only expressed by *L. pneumophila* when the microorganism is within a host cell.

We hope that these two strategies will allow identification of most of the factors encoded by *L. pneumophila* that mediate the striking rearrangement of host cell organelles and facilitate growth of the microorganism.



Scanning electron micrograph of lucerifase reporter mycobacteriophages (LRMs) adsorbed to the surface of Mycobacterium bovis*—BCG—cells. This is a model for the use of LRMs for rapid detection of* Mycobacterium tuberculosis *cells in a human clinical sample. The phage particle, once adsorbed, injects its genome, into which the firefly luciferase gene has been inserted. Expression of the gene causes the mycobacterial cells to emit light, allowing sensitive and rapid detection of* M. tuberculosis *cells and assessment of their drug-susceptibility patterns.*

Scanning photograph by Rupa Udani and Frank Macalusa in the laboratory of William Jacobs.

Genetic Approaches to the Control of Mycobacterial Diseases

William R. Jacobs, Jr., Ph.D.—Assistant Investigator

Dr. Jacobs is also Associate Professor of Microbiology and Immunology and of Molecular Genetics at Albert Einstein College of Medicine. He received a B.A. degree in mathematics at Edinboro University of Pennsylvania and a Ph.D. degree in molecular cell biology from the University of Alabama at Birmingham. His doctoral work on Mycobacterium leprae *was performed in the laboratory of Josephine Clark-Curtiss and Roy Curtiss III, first at the University of Alabama and then at Washington University, St. Louis. His postdoctoral studies with Barry Bloom focused on developing systems to express foreign antigens in the tuberculosis vaccine strain BCG (bacille Calmette-Guérin).*

TUBERCULOSIS, caused by *Mycobacterium tuberculosis*, continues to be a serious health problem throughout the world. Even in the developed countries, this plague has been escalating in recent years. The World Health Organization (WHO) estimates a worldwide incidence of approximately 8 million new cases and over 3 million deaths annually. After 32 years of steady decline in the United States, the number of new cases has taken a surprising and alarming upward turn for each of the last five years. In 1990 there was a 34 percent increase in New York City and a 10 percent increase nationally. More recently a strain of multidrug-resistant *M. tuberculosis* has resulted in numerous deaths, demonstrating the dire reality that tuberculosis caused by a virtually invincible pathogen is now spreading.

Other mycobacteria play significant roles in world health. *Mycobacterium leprae* is the causative agent of leprosy, an affliction know to the ancients that affects over 13 million in the world today. Another agent is *Mycobacterium avium*, a major opportunistic pathogen for many individuals with AIDS.

Our ongoing studies are aimed at developing novel ways of treating mycobacterial disease, particularly tuberculosis, using molecular genetic approaches. In addition, we have taken a new look at the vaccine known as BCG (bacille Calmette-Guérin), which has been used widely since 1924 to prevent tuberculosis. We are genetically engineering BCG into a multivalent vaccine vector that can elicit a protective immune response to a wide variety of bacterial, viral, and parasitic pathogens. The development of systems to alter mycobacteria genetically should permit both goals to be achieved.

Genetic manipulation of these microorganisms has only been possible in the last few years. *M. tuberculosis*, like other mycobacteria, is difficult to analyze genetically for a variety of reasons. First, the tubercle bacillus, which multiplies only once every 24 hours, requires 3 weeks to form a colony from a single cell. In contrast, *Escherichia coli* yields visible colonies in 8 hours. The leprosy bacillus has yet to be cultivated in the laboratory and can only be grown in mouse footpads or the nine-banded armadillo. Over the last five years, we have developed a series of phage- and plasmid-based vectors that have enabled the efficient introduction of recombinant DNA into mycobacteria. Recombinant DNA technologies have opened exciting new doors to basic knowledge about these organisms and the ways they cause infection. In addition, these technologies offer novel reagents, such as luciferase reporter phages, and recombinant vaccines that could play major roles in combating human disease.

Epidemiological Analysis of Tuberculosis Infections

Use of restriction fragment length polymorphisms (RFLPs) can play a key role in determining the mode of transmission of tuberculosis, as individual isolates can be tracked from one infected person to the next. The recent increases in the incidence of tuberculosis in the United States seem to be largely associated with the AIDS epidemic. However, it was unclear whether the tuberculosis seen in AIDS patients results from reactivated disease, reflecting pre-AIDS exposure, or from a first-time infection. In collaborations with researchers in San Francisco, including Gary Schoolnik (HHMI, Stanford University), we have undertaken analyses to distinguish these two possibilities.

Different isolates of *M. tuberculosis* have different RFLP patterns when probed with a particular DNA element found in *M. tuberculosis* strains. In isolates from a recent tuberculosis outbreak among AIDS patients in a group home, RFLP analysis revealed that 11 persons had all been infected with the same strain of *M. tuberculosis*. This demonstrates that AIDS patients are highly susceptible to *M. tuberculosis* infection and confirms that tuberculosis is highly contagious. These clear results should be translatable into better public health care policies. Similar analy-

ses are under way to track *M. tuberculosis* isolates that are multiply drug resistant.

Luciferase Reporter Phages

Accurate diagnosis of *M. tuberculosis* infection routinely requires 4–6 weeks to allow cultivation of the tubercle bacilli and time to perform the appropriate tests. Assessment of the drug-susceptibility patterns of clinical isolates can require another 4–6 weeks of analysis. In light of the increasing numbers of drug-resistant specimens, we set out to devise a novel diagnostic test that would shorten the time required for both procedures. The test uses a mycobacteriophage, a virus that infects *M. tuberculosis*, into which we have cloned a gene that encodes a reporter enzyme, the luciferase that makes fireflies glow. Luciferase reporter mycobacteriophage (LRM) particles, when mixed with bacterial cells, result in the production of light.

The LRM test is extraordinarily sensitive, as photons can be detected at extremely low levels. In addition, it is exquisitely specific, as the phage only attaches to mycobacterial cells. Thus the production of light reveals the presence of *M. tuberculosis* in the sample. We have undertaken studies in clinical application. Since the luciferase reaction requires ATP for photon production, we also hope to use the test to assess the metabolic activity of a cell. And since drug treatment of mycobacterial cells abrogates this activity, we are also exploring the use of the test to distinguish drug-resistant from drug-sensitive cells.

Genetic Analysis of Mycobacterial Virulence Determinants

We would like to know why pathogens such as *M. tuberculosis* or *M. leprae* cause severe disease when BCG can elicit an effective immune response against these organisms. To approach these questions, we have focused on developing systems for defining and characterizing the genes of mycobacteria.

The first approach is to generate specific insertion mutations within the virulent *M. tuberculosis* genome and to screen for mutants that are no longer virulent in our animal models. Toward this goal, we are doing a number of studies designed to generate random insertions within the genome by taking advantage of natural mycobacterial transposons or recombination systems. Complementary to this mutant isolation is the development of both extrachromosomal and integrating vectors that permit highly efficient introduction of libraries of genes from virulent mycobacteria into avirulent strains.

The combination of these two strategies has already allowed us to identify genes necessary for the biosynthesis of amino acids, purines, and complex polysaccharides found on the surface of the pathogenic mycobacteria. We are currently developing a variety of systems to screen these libraries for genes that confer virulence characteristics to the avirulent organisms. By identifying the genes and their products responsible for specific virulence phenotypes, we hope to make possible the design of strategies to prevent or control mycobacterial disease. In addition, the genetically engineered avirulent mutants should provide novel vaccine candidates.

Recombinant BCG Vaccines and Novel Vaccine Strategies

BCG is an attenuated mutant of bovine tuberculosis bacillus. It has been used as a vaccine against tuberculosis in humans for over 50 years. The bacterium possesses several unique properties that make it ideal for use as a live vector in generating multivalent vaccines. BCG is a safe vaccine, having been used in 2.5 billion individuals with a mortality rate significantly lower than that of the smallpox vaccine. It is the only live vaccine other than oral polio that WHO recommends for use in infants. Also, the mycobacterial cell wall has potent adjuvant properties that can engender excellent humoral responses; and since BCG is stored within macrophages, which are key antigen-presenting cells, it follows that it can also elicit cellular immune responses.

In the last few years, our laboratory, in collaboration with Barry Bloom (HHMI, Albert Einstein College of Medicine), MedImmune, and Graham Hatfull at the University of Pittsburgh, has developed a series of expression vectors and transformation systems whereby foreign genes encoding antigens from virtually any pathogen can be cloned and expressed in BCG. Mice immunized with these recombinant BCG cells have been shown to elicit both humoral and cellular immune responses to the expressed foreign proteins. We are currently cloning and expressing genes from pathogens that cause leishmaniasis, schistosomiasis, and toxoplasmosis in BCG. Immunization of mice with recombinant BCG will provide useful models to test the types of immune responses that can be elicited with these live bacteria. Ultimately we hope to engineer a recombinant BCG that could protect humans from these dread diseases.

Mechanisms of Neurotransmitter Storage and Release

Reinhard Jahn, Ph.D.—Associate Investigator

Dr. Jahn is also Associate Professor of Pharmacology at Yale University School of Medicine. He received his Ph.D. degree at the University of Göttingen, Germany. His postdoctoral training was with Hans Dieter Söling at the University of Göttingen and later with Paul Greengard at Yale University and the Rockefeller University. He was Assistant Professor at the Rockefeller University and subsequently headed a research group at the Max Planck Institute for Psychiatry, Martinsried.

NERVE cells, or neurons, communicate with each other and with other cells by means of small molecules, the neurotransmitters. This communication occurs at specialized contact zones, the synapses. Upon arrival of incoming action potentials, the sender cell releases its neurotransmitters, which cross the synaptic cleft. The plasma membrane of the receiving cell has specific receptor molecules that, in turn, translate the signals into functional changes.

In the resting state, the presynaptic nerve endings of the sender cell store their neurotransmitters in small membrane-enclosed compartments, the synaptic vesicles. When an action potential arrives, voltage-gated calcium channels open and calcium ions (Ca^{2+}) enter the terminal from the extracellular space. Within a fraction of a millisecond, synaptic vesicles fuse with the plasma membrane, releasing their contents. The vesicle membrane protein is then retrieved by endocytosis for use in regenerating fusion-competent synaptic vesicles.

The details of this membrane recycling are still largely unclear, and the enzymes catalyzing the major steps have not been identified. As a starting point for detailed functional analysis, we and others have characterized the major membrane proteins of synaptic vesicles. This was facilitated by the fact that synaptic vesicles are abundant and can easily be purified in large amounts. Due to the smallness of the vesicles, the number of protein species per individual vesicle is inherently limited. Therefore it should be possible to identify most, if not all, of the major protein constituents. To date, several families of unique vesicle proteins have been characterized, and advanced tools for their study are available. We are currently involved in a systematic chemical analysis of the vesicle membrane to identify the remaining proteins of this organelle.

Parallel to this protein analysis, we have recently begun to establish assay systems for individual steps of the vesicle cycle. One of the model systems accessible for biochemical analysis is the isolated nerve terminal. Upon homogenization of nervous tissue, nerve terminals, though sheared off their axons, reseal and remain functional for several hours after isolation. Application of depolarizing stimuli or Ca^{2+} ionophores causes a massive exocytotic transmitter release and a parallel increase in membrane turnover.

We found that during exo-endocytosis, one of the vesicle proteins, the GTP-binding protein rab3A, dissociates from the vesicle membrane and reassociates again at a later stage of the membrane cycle. This dissociation-association cycle is probably necessary for an orderly and sequential processing of the vesicle membrane. Thus small GTP-binding proteins may serve in the nerve terminal, as in other organelles, as "status indicators" for the recycling membrane, being associated only with a specific step in the life cycle of the vesicle membrane. The biochemical events leading to GTPase activation and membrane dissociation and reassociation are presently under study. In addition, the individual G proteins should allow an isolation of the compartments representing separate steps in the membrane cycle.

Thus we have isolated clathrin-coated vesicles from nerve terminals by conventional procedures and have analyzed their membrane and coat composition. We have also developed immunoisolation procedures for synaptic vesicle-derived membrane populations, using monoclonal antibodies directed against individual membrane components. We hope these studies will aid in the functional definition of the compartments involved in synaptic vesicle recycling and will form the basis for the reconstitution of individual steps of the membrane cycle in cell-free systems.

Molecular Studies of Voltage-Sensitive Potassium Channels

Lily Y. Jan, Ph.D.—Investigator

Dr. Jan is also Professor of Physiology and Biochemistry at the University of California, San Francisco. During her graduate study at the California Institute of Technology with Jean Paul Revel and Max Delbrück, Dr. Jan localized the visual pigment rhodopsin at the ultrastructural level. Because she studied high-energy theoretical physics before becoming a biology student, her Ph.D. degree was in physics and biophysics. She stayed at CalTech to do postdoctoral research with Seymour Benzer and began to collaborate with her husband, Yuh Nung Jan. Their first collaboration resulted in the identification of the Shaker *locus as a potential structural gene for a potassium channel. Before accepting faculty appointments at UCSF, the Jans worked in Stephen Kuffler's laboratory at Harvard Medical School.*

VOLTAGE-sensitive potassium channels probably constitute the most diverse and widespread class of ion channels. More than 30 different types of potassium channels have been characterized. They differ in their voltage sensitivity, their kinetic properties, and their sensitivity to second messengers within the cell. Potassium channels have been found in almost every eukaryotic cell type examined, in both the animal and the plant kingdoms. They are important for a wide range of physiological functions, including insulin release due to raised glucose levels, proliferation of lymphocytes induced by mitogens, and the movements of leaflets in plants or the opening and closing of leaf stomatal pores. In the mammalian nervous system, potassium channels control excitability and the strength of signaling between nerve cells. Indeed, some of the potassium channels have been implicated as playing a role in learning and memory.

To study how the diversity of potassium channels arises and how they serve the wide variety of cellular functions, one needs to study these channels biochemically as well as biophysically. However, they are difficult to purify because they are rather heterogeneous and inaccessible. For this reason we have taken advantage of the well-developed genetic technologies applicable to the fruit fly *Drosophila melanogaster*. In this organism, if a gene (say, coding for a potassium channel) can be identified by the abnormalities caused by its mutations, one can clone it for molecular studies of the gene product.

More than a decade ago, Yuh Nung Jan (HHMI, University of California, San Francisco), Mike Dennis, and I found that mutations at the *Shaker* locus cause prolonged transmitter release from the motor nerve terminal, probably because of a defect in potassium channel function. The locus was subsequently cloned by Diane Papazian, Tom Schwarz, and Bruce Tempel in our laboratory. It codes for proteins that contain multiple stretches of hydrophobic amino acids that can potentially span the cell membrane; several proteins that differ in sequence flanking these stretches of hydrophobic amino acids are generated because of alternative splicing of the primary transcript. Leslie Timpe demonstrated that RNA encoding for these *Shaker* proteins of known deduced sequence, when injected into frog oocytes, causes the functional expression of potassium channels of different kinetic properties. Since the different *Shaker* proteins are present in different regions of the fly brain, they are likely to give rise to different subtypes of A channels.

Starting with the *Shaker* gene in the fruit fly, Bruce Tempel isolated a gene in the mouse that codes for a potassium channel in the mouse brain. This protein is 65 percent identical in its sequence to the *Drosophila* potassium channel protein. In frog oocytes, the gene produces potassium channels that do not inactivate rapidly. Analysis of the distribution of amino acid residues that appear to be essential and have been totally conserved over 600 million years has provided some clues to the channel's structure. Now over 15 different mammalian potassium channel genes have been characterized by a number of laboratories. One of these (the rat *Shal1* gene), cloned and characterized by Tim Baldwin, Meei-Ling Tsaur, and George Lopez in our laboratory, produces a rapidly inactivating potassium channel in frog oocytes and is expressed in the heart as well as the brain. The high degree of conservation among these mammalian channels and the fruit fly potassium channels reiterates the point that any experimental organism, as long as it is amenable to the specific type of experimentation, will reveal information of medical interest.

Having cloned some of the potassium channel genes, we can now ask how this channel works. How does it detect a voltage change across the cell membrane and, responding, open? How does it "inactivate" after it opens? How does it discriminate between sodium and potassium ions and show exquisite selectivity? To probe these questions, we have altered specific residues of the potassium channel to see how the various functions are affected. Studies in our laboratory (Ehud

Isacoff, George Lopez, Diane Papazian, and Leslie Timpe; this work is supported by a grant from the National Institutes of Health) and others have provided evidence for the involvement of specific structural elements in the detection of voltage changes across the cell membrane and in the subsequent conformational changes that open the channel and allow the cytoplasmic mouth of the pore to interact with the inactivation gate, thereby terminating channel opening.

For studies of the biological functions of potassium channels, we have chosen to concentrate on the mammalian heart and hippocampus. A variety of cardiac potassium channels have been characterized biophysically and are important in controlling the rhythmic heartbeat. Molecular studies of these channels will not only contribute to our understanding of channel function but will also be relevant clinically, for example, in the development of more-specific drugs for cardiac arrhythmia. The hippocampus is a region of the mammalian brain that appears to play an important role in learning and memory. It has also been studied extensively in experimental paradigms that induce epileptic activity. By cloning and analyzing potassium channel genes that are expressed in this tissue, we hope to learn about the involvement of these potassium channels in the normal function and pathology of the nervous system.

Using specific probes for individual potassium channel polypeptides and their messenger RNA, Morgan Sheng and Meei-ling Tsaur have found distinct patterns of expression in the mammalian brain. In addition to spatial regulation at the level of brain regions as well as specific neuronal types, dynamic changes in potassium channel gene expression have also been observed in the adult brain following pentylenetetrazole-induced seizure. (This work is supported by a grant from the National Institute of Mental Health.) The observed decrease of specific potassium channel transcripts in the excitatory neurons in the dentate gyrus of the hippocampus is likely to lead to an increase of excitability. These observations suggest that potassium channel gene regulation may contribute to long-term neuronal plasticity.

Neural Development in Drosophila

Yuh Nung Jan, Ph.D.—Investigator

Dr. Jan is also Professor of Physiology and Biochemistry at the University of California, San Francisco. Although Dr. Jan went to the California Institute of Technology to study theoretical physics, he instead became interested in biology and received his Ph.D. degree from Caltech in biophysics and physics. While there he studied sensory transduction of the fungus Phycomyces *with Max Delbrück. Dr. Jan began his study of the nervous system during postdoctoral research with Seymour Benzer at CalTech and continued this line of research with Stephen Kuffler at Harvard Medical School. His primary interest remains the nervous system.*

HOW a nervous system is organized during development is a major unresolved problem in biology. For several years our laboratory has been interested in the following questions: How do neurons arise from undifferentiated ectodermal cells? What gives the neurons their individual identity in terms of shape and function?

Our long-term goal is to understand these processes at the molecular level, and our approach is a genetic one. We first isolate mutations that affect neurogenesis, neuronal type, or axonal pathway formation. Identification of these mutations can lead to the isolation of important genes.

During the last few years, our laboratory has been engaged in an extensive search and analysis of mutants affecting neural development in the fruit fly *Drosophila*. To identify and analyze such mutants, we have been using the embryonic sensory nervous system, which has been characterized in considerable detail at the single-cell level. Roughly half of the *Drosophila* genome has been screened for mutations that alter the peripheral nervous system (PNS), resulting in the identification of a number of genes that specify cell fate in the embryo. Analysis of those genes had led us to propose a "progressive determination of the PNS" model.

Early during embryogenesis, cells in different locations within the ectodermal layer acquire unequal developmental potential as a result of the actions of the "prepattern genes." These include genes that specify dorsoventral and anteroposterior orientation as well as segmentation. Cells from some of the domains are affected by the action of "proneural" genes, which apparently endow cells with the competence to become neuronal precursors. Genes of the *achaete-scute* complex (*AS-C*) and *daughterless* (*da*) belong to this group. Both *AS-C* and *da* encode proteins with the helix-loop-helix motif. It is likely that products of these genes may form homo- or heterodimers that bind to DNA and regulate the transcription of target genes in order to initiate neuronal precursor development.

As a neuronal precursor forms, it inhibits neighboring cells from doing so. This "lateral inhibition" involves the action of six known "neurogenic" genes. Removing the function of any of the six, *Notch*, *Delta* (*Dl*), the *Enhancer of split* complex [*E(spl)-C*], *mastermind* (*mam*), *neuralized* (*neu*), and *big brain* (*bib*), leads to hypertrophy of both the central nervous system (CNS) and the PNS, presumably as a result of losing lateral inhibition.

There appear to be at least two independent cell-cell interaction pathways. One is mediated by the gene products of *Notch* and *Dl*, both coding for membrane proteins with epidermal growth factor (EGF)-like repeats. *E(spl)-C*, *mam*, and perhaps *neu* are involved in this pathway. The second pathway is mediated by *bib*, which encodes a membrane protein with significant homology to the bovine major intrinsic protein (MIP), soybean nodulin 26, and *Escherichia coli* glycerol facilitator, which allows passive transport of small molecules such as glycine.

The commitment of neuronal precursors may involve the actions of a group of "master regulatory" genes, which endow the precursor with certain unique properties of the nervous system. The identity of a neuronal precursor is further specified by "neuronal-type selector" genes. For example, the *cut* locus is required for external sensory organs to acquire their correct identity. In the absence of *cut* function, these organs are transformed into chordotonal organs. Normally *cut* is expressed in sensory organ precursors but not in chordotonal organ precursors, and we think that the *cut* gene determines which organ the precursor will develop into. The *cut* product contains a homeodomain and is likely to act as a transcription factor regulating the expression of downstream differentiation genes.

Recently we began to work on the problem of axonal pathfinding and target recognition in *Drosophila*. Tracing the axonal pathway in the fly nervous system had been problematic in the past. Because of the small size of the neurons, it was difficult to use traditional methods to trace pathways, such as filling neurons with fluorescent dyes and other tracers. However, Ed Giniger, a

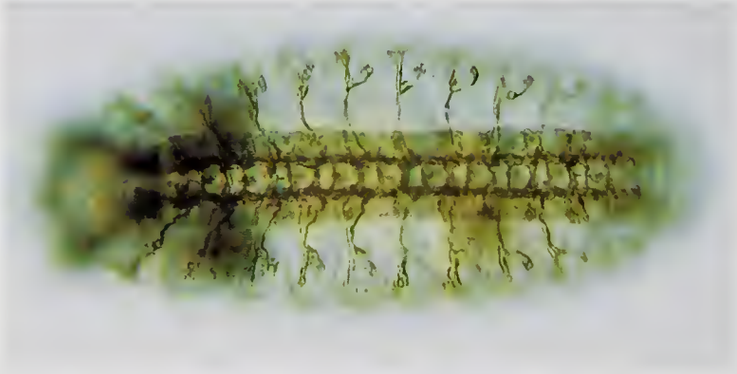
postdoctoral fellow in our laboratory, found a solution to this difficulty.

He modified the enhancer trap method we used previously by replacing the β-galactosidase (β-gal) reporter gene with a kinesin–β-gal fusion gene. The idea is that the kinesin part of the fusion protein will drag β-gal down the axon to the terminal. The idea works very nicely. We have now generated over a thousand enhancer trap lines with kinesin-*lacZ* as a reporter gene and, in more than a hundred of those, have labeled subsets of CNS and/or PNS pathways, including sensory and/or motor projections. We would like to choose a small number of lines that reveal simple axonal pathways (e.g., sensory neuron or motor neuron pathways) as a basis for a mutant screen.

Based on sequence information of a number of genes involved in neural development, it seems that the majority of these genes contain a previously identified functional motif—e.g., the EGF repeat found in *Notch*, the tyrosine kinase domain in *sevenless*, the homeodomain in *cut*, and the helix-loop-helix motif in *da* and *AS-C*. In each case the existence of such a motif immediately provides strong suggestions for the likely mode of gene action, which may be tested experimentally.

The existence of readily identifiable motifs in the majority of genes involved in neural development suggests that cells have a limited repertoire of mechanisms for essential regulatory functions. These include various signaling systems, such as ligands, receptors, second messenger systems, and regulators of gene expression. Perhaps only a modest number of new tools had to be invented for the formation of the nervous system. Many of the cellular mechanisms used in neural development may already have been available before the nervous system evolved. Understanding neural development may require an understanding of the usage and manipulations of these basic functional motifs.

It is also apparent that many functional motifs have been highly conserved during evolution. Several hundred million years separate the insects and vertebrates, yet the aforementioned motifs (the EGF repeat, tyrosine kinase domain, homeodomain, helix-loop-helix) are common to both and remarkably enduring. Thus studying development in the simpler organisms with well-developed genetics, such as the fruit fly and the nematode, may provide valuable background for the investigation of corresponding problems in higher forms.



To follow the development of a nerve cell, the neurobiologist will typically inject it with a tracer dye. In view of the minuteness of the fruit fly embryo, however, a method was developed to cause the fly to fill its own neurons with a marker. A bacterial enzyme was targeted to nerve processes by fusing the gene for kinesin, a molecular motor, to the gene encoding β-galactosidase, an enzyme for which there are simple colorimetric assays, and Drosophila *stocks were established that express the fusion protein. The example shown here displays the pattern seen when the protein is made in all the neurons of the embryo. The central ladder-like structure comprises the axons of the ventral nerve cord, and the fine lateral extensions are the peripheral neurons and axons.*

Research and photograph by Ed Giniger in the laboratory of Yuh Nung Jan.

Activation of CD4 T Cells

Charles A. Janeway, Jr., M.D.—Investigator

Dr. Janeway is also Professor of Immunobiology at Yale University School of Medicine. He is a graduate of Harvard College (B.A., chemistry) and of Harvard Medical School. He had research training during medical school at the National Institute for Medical Research in London, England; postdoctoral training in internal medicine at the Peter Bent Brigham Hospital in Boston; and immunology research experience at NIH and the Biomedical Center in Uppsala, Sweden. Dr. Janeway was awarded the degree of Doctor Honoris Causa by the Copernicus Medical School of Jagellonian University in Cracow, Poland.

THE critical event in most immune responses is the activation of specific CD4 T lymphocytes. When these cells are lost, as happens in AIDS, the immune system can no longer defend the host against infection. Our laboratory is studying the mechanism by which these cells become active. This is a complex process involving changes in both the cell that presents the activating antigen and the T cell itself.

The process requires two events. First, the antigen-specific receptor on the T cell must encounter its appropriate ligand on the surface of an antigen-presenting cell. This results in specific signaling of the T cell that will cause it to become immunologically inert, or anergic. Anergic T cells cannot mount responses. In order for antigen recognition to lead to an effective response, the CD4 T cell must receive a second signal or co-stimulator. Studies in our laboratory have characterized the chemical structure of several peptides that are bound by class II molecules of the major histocompatibility complex (MHC) and thus recognized by CD4 T cells. Second, we have identified two molecules expressed on antigen-presenting cells that are capable of co-stimulating the activation of resting, naive CD4 T cells. Finally, we have shown that a single cell must present both the peptide-MHC complex and the co-stimulator in order for normal CD4 T cells to be activated effectively.

The ligand recognized by the receptor on CD4 T cells consists of a peptide fragment of a foreign protein antigen bound to a class II MHC-encoded molecule. Our laboratory has isolated peptides from such molecules and characterized their structure. These molecules are approximately 15 amino acids in length and have specific residues at certain positions for each different MHC class II molecule. Although the cleavage of these fragments from the native proteins has not been characterized, the results of our analysis suggest that the MHC molecule protects the peptide fragment from further proteolytic degradation.

A striking result in these studies was the finding that 12 percent of normal MHC class II molecules on antigen-presenting cells were in complex with a particular self peptide. As a number of similarly dominant peptides were observed, it seems possible that the total complexity of self peptides to which the immune system must be tolerant may be quite limited. This is important, as it allows the maximum complexity of foreign peptides to be recognized.

Of even greater interest is the finding that this specific complex is found richly represented on cells in the thymic medulla, including bone marrow–derived cells, but is present only at very low levels on thymic cortical epithelium. These two tissues are known to carry out distinctive events involving self MHC-peptide complexes in the maturation of T cells within the thymus. The finding that a self peptide is unevenly distributed between these cell types suggests that T cell development occurs under different selecting environments for different processes. The antibody recognizing this processed self peptide bound to self MHC class II will allow our laboratory to probe this process in living animals. The same technology is now being applied to the analysis of autoantigens and peptides generated by intracellular infectious agents such as *Salmonella typhimurium*.

In addition to ligation of the receptor, the same antigen-presenting cell must also express co-stimulatory molecules on its surface in order to activate CD4 T cells optimally. We have confirmed the importance of the B7 molecule, a known co-stimulator of human T cells, in the activation of murine CD4 T cells and have discovered that the heat-stable antigen can also serve as a co-stimulator for murine T cells. Either of these proteins, transfected into fibroblasts, can confer their CD4 T cell–activating properties. Moreover, the two proteins act in a strongly synergistic manner on normal antigen-presenting cells to promote the proliferation of CD4 T cells.

Finally, we have shown that B7 expression is regulated in parallel with co-stimulatory activity when B lymphocytes respond to a variety of microbial products. This regulated behavior of co-stimulators allows B cells to present antigen to CD4 T cells for tolerance in the absence of micro-

bial infection or for immunity in its presence. We documented the importance of this by showing that tolerance to a self protein could be broken by inducing the expression of co-stimulator molecules on self antigen-reactive B cells. This last system may help to explain the role of specific pathogens in the initiation of autoimmunity. Parts of this work are also supported by grants from the National Institutes of Health.

The T cell perceives the complex of foreign peptide and self MHC class II molecule with a receptor resembling the antigen-binding fragment of an immunoglobulin molecule. Although the most variable portion of this receptor lies in its central region, generated by the joining of several different gene segments, some variability is also expressed around the periphery of the ligand-binding site. This raises the question of what residues on the peptide-MHC complex are contacted by these peripheral antigen-binding loops on the T cell receptor.

We recently showed that at least one of these loops makes contact with one side of the MHC molecule binding the foreign peptide. Detailed mutagenesis studies of both the T cell receptor and the MHC protein should soon reveal the definitive orientation of the T cell receptor to its ligand.

The information derived from our studies on T cell activation is being applied to the analysis of a model autoimmune disease, the insulin-dependent diabetes mellitus that occurs spontaneously in non-obese diabetic mice. We have isolated cloned T cells capable of invading the islets of irradiated mice and destroying β-cells, producing diabetes. In addition, we have identified other cells that appear to protect the islet. Analysis of the peptide-MHC ligands recognized by these various cell types, and of the events involved in their activation, should permit a better understanding of diabetes mellitus and lead to its specific immunomodulation. This work is also supported by a grant from the National Institutes of Health. Thus the basic science base of the laboratory is being applied to disease models in hopes of improving therapy in immunological diseases.

Control of Cell Pattern in the Developing Nervous System

Thomas M. Jessell, Ph.D.—Investigator

Dr. Jessell is also Professor of Biochemistry and Molecular Biophysics at Columbia University College of Physicians and Surgeons and a member of the Center for Neurobiology and Behavior. He received his Ph.D. degree in neurobiology from Cambridge University, England, and was elected a research fellow of Cambridge's Trinity College. He was a postdoctoral fellow in Gerald Fischbach's laboratory at Harvard Medical School. Next he served as Assistant Professor of Neurobiology at Harvard Medical School, before moving to Columbia University.

OUR research is aimed at understanding the mechanisms that control cell patterning within the developing vertebrate nervous system. The major focus is on deciphering how discrete neural cell types appear at defined positions in the embryo. In addition, we are examining the roles of diffusible and cell surface molecules in the guidance of developing axons in the spinal cord. Our studies over the past several years have provided evidence that the floor plate, a specialized group of neuroepithelial cells, has critical roles in both the control of cell identity and in axon guidance. Within the past year, further details of the functions and molecular properties of the floor plate have become apparent.

Control of Motor Neuron Differentiation

Our previous studies had shown that signals originating from the floor plate regulate the identity of specific cell types within the neural tube. In order to examine in more detail the actions of the floor plate on neural cell patterning, we have begun to focus on one well-characterized neuron class, the spinal motor neuron. There is extensive information on the mechanisms that control the pathfinding of motor axons and the formation of synapses at the neuromuscular junction, but the events that control the generation of motor neurons remain largely obscure.

Insight into the molecular mechanisms involved in the generation of motor neurons by floor plate–derived signals requires the identification of genes that are expressed at the initial stages of motor neuron differentiation. In collaboration with Thomas Edlund's laboratory in Umea, Sweden, we have found that embryonic chick motor neurons express a homeobox gene called *Islet-1* (*Isl-1*). This is a member of the subfamily of homeobox genes that contain cysteine-rich regions called LIM domains. Other members of the family include *Lin-11* and *Mec-3*, which have been shown to regulate cell fate in *Caenorhabditis elegans*.

Isl-1 binds to enhancer elements in the rat insulin gene and is expressed in pancreatic islet cells and in a subset of neurons, including motor neurons. In the embryonic chick spinal cord, *Isl-1* immunoreactivity is first detected in the nuclei of cells in the ventral region, lateral to the floor plate. The number of *Isl-1*$^+$ cells in the ventral spinal cord increases markedly during early spinal cord development. We established that the ventral spinal cord cells that express *Isl-1* are motor neurons by retrograde injection of horseradish peroxidase (HRP) into motor axons in the ventral root.

Analysis of the expression of motor neuron markers by chick spinal cord cells *in vivo* has provided evidence that the differentiation of motor neurons is dependent on inductive signals from the floor plate. In agreement with this, the spinal cord of embryos that had received floor plate grafts contained additional ectopic *Isl-1*$^+$ cells. The induced *Isl-1*$^+$ cells also expressed the SC1 glycoprotein and sent axons out of the spinal cord consistent with their identity as motor neurons.

These results provide evidence that signals from the floor plate can induce the expression of *Isl-1* in dorsal neural tube cells. Elimination of the notochord and floor plate before neural tube closure results in the development of a spinal cord devoid of ventral neuronal types, including motor neurons. *Isl-1* expression in the ventral spinal cord is also dependent on signals from the notochord and floor plate. These results support the idea that elimination of the notochord and floor plate prevents the initial steps in the differentiation of motor neurons.

Homeobox genes are involved in many aspects of vertebrate development. The expression pattern of *Hox* genes along the anteroposterior axis of the neural tube, and of *Pax* genes along the dorsoventral axis, together with the phenotypes that result from inactivation of some of these genes, suggests that they contribute to the regional patterning of the developing nervous system. In contrast, the restricted expression of *Isl-1* and the involvement of related LIM-homeodomain proteins in the determination of cell fate in *C. elegans* suggest that *Isl-1* may be involved in specifying the fate of specific neuronal subtypes —in particular, motor neurons.

Floor Plate–Specific Genes in Axon Guidance

After the identity of motor neurons and other spinal cord neurons has been established, the floor plate appears to provide both long-range and local guidance cues that promote the growth of axons to and across the ventral midline of the spinal cord. First, the floor plate secretes a diffusible chemoattractant that can orient the growth of axons of commissural neurons *in vitro* and may account for the homing of these axons to the floor plate *in vivo*. Second, the floor plate may contribute to the change in trajectory of commissural axons from the transverse to the longitudinal plane that occurs immediately after crossing the ventral midline. In support of this proposal, genetic mutations in mice and zebra fish that result in the absence of the floor plate during embryonic development lead to errors in the pathfinding of commissural axons at the midline of the spinal cord. Third, the floor plate may promote the fasciculation of commissural axons that occurs after they cross the midline of the spinal cord by regulating the expression of glycoproteins of the immunoglobulin superfamily. The specialized role of the floor plate in vertebrate neural development has parallels in invertebrate organisms, in that cells at the midline of the embryonic *Drosophila* and *C. elegans* central nervous systems have been implicated in neural patterning and axon guidance.

To identify molecules that may mediate the diverse functions of the floor plate during early neural development, we have used subtractive hybridization techniques to isolate cDNA clones expressed selectively by the floor plate. One of these cDNA clones encodes a novel secreted protein, F-spondin, which is expressed at high levels by the rat floor plate during embryonic development. F-spondin contains domains similar to those present in thrombospondin and other proteins implicated in cell adhesion and neurite outgrowth. *In vitro* assays show that F-spondin promotes neural cell adhesion and neurite outgrowth, suggesting that its secretion by the floor plate contributes to the growth and guidance of axons in the developing central nervous system.

The F-spondin protein may be associated with the extracellular matrix, since it has several clusters of basic residues that function as glycosaminoglycan-binding domains in other secreted proteins. The restricted distribution of F-spondin mRNA in the embryonic nervous system contrasts with the distribution of other secreted glycoproteins that promote neural cell adhesion and neurite outgrowth. For example, the expression of F-spondin mRNA is more restricted than that of thrombospondin and of tenascin, which appear to be expressed widely in the embryonic central nervous system.

The prominent expression of F-spondin in the floor plate suggests that the protein may be involved in the plate's development or functions. Midline neural plate cells that give rise to the floor plate undergo marked changes in cell shape during the closure of the neural tube. Thus one possible function of F-spondin could be to mediate adhesive plate cell interactions that maintain the integrity of the floor plate during formation of the embryonic spinal cord.

The expression of F-spondin mRNA in floor plate cells is highest at the time of the plate's suggested role in the chemotropic guidance of commissural axons. However, recombinant F-spondin does not mimic the ability of the floor plate–derived chemoattractant to promote the outgrowth of commissural axons from dorsal spinal cord explants. This suggests that F-spondin may not be involved in the long-range guidance of commissural axons to the floor plate, at least through chemotropism.

F-spondin is more likely to be involved in the contact-dependent guidance of commissural axons once they reach the ventral midline of the spinal cord under the influence of distinct chemotropic guidance cues. The growth cones of commissural neurons cross the midline by growing between the basal surface of floor plate cells and the underlying basal lamina. Floor plate–secreted F-spondin may accumulate at high levels in association with the basal surface of floor plate cells or with the underlying basal lamina, thus generating a difference in adhesive properties of the floor plate and the lateral neuroepithelium. The growth cones of commissural neurons may adhere preferentially to F-spondin, prompting them to change trajectory at the boundary of the floor plate and lateral neuroepithelium. It is also possible that F-spondin has a more active signaling role that induces changes in the properties of commissural growth cones, permitting them to respond to other midline guidance cues.

Energy-transducing Membrane Proteins

H. Ronald Kaback, M.D.—Investigator

Dr. Kaback is also Professor of Physiology and of Microbiology and Molecular Genetics in the Molecular Biology Institute of the University of California, Los Angeles. He received his M.D. degree from the Albert Einstein College of Medicine, interned at Bronx Municipal Hospital Center, and did postdoctoral research in physiology at Einstein. Subsequently he conducted research in membrane biochemistry at the National Heart Institute and the Roche Institute of Molecular Biology, chairing at Roche the Department of Biochemistry. Dr. Kaback is a member of the National Academy of Sciences and the American Academy of Arts and Sciences. Among his honors is the Distinguished Alumnus Award from the Albert Einstein College of Medicine.

THE molecular mechanism of energy transduction in the membranes of living cells is an enigma. Although the immediate driving force for many seemingly unrelated processes, such as active transport, oxidative phosphorylation, and bacterial motility, is a bulk-phase, transmembrane electrochemical H^+ or Na^+ gradient, the molecular mechanism(s) by which free energy stored in these gradients is transduced into work or into other forms of chemical energy (e.g., ATP) remains unknown. In order to gain insight into this important basic problem, studies in our laboratory have focused on an enzyme, the lactose (lac) permease of the bacterium *Escherichia coli*, as a paradigm.

The ability of *E. coli* to accumulate the sugar lactose and other β-galactosides against a concentration gradient is dependent upon lac permease, a very hydrophobic plasma membrane protein that catalyzes the coupled translocation of a single sugar molecule with a single H^+ (i.e., symport or co-transport). Under physiological conditions, lac permease utilizes free energy derived from downhill translocation of H^+ to drive accumulated β-galactosides against a concentration gradient or, conversely, uses free energy released from downhill translocation of β-galactosides to drive uphill translocation of H^+. The polarity of the reaction reflects the direction of the concentration gradient of the substrate. As such, lac permease represents a huge number of machines that catalyze similar reactions in virtually all biological membranes, from archebacteria to the mammalian central nervous system.

The permease is encoded by the *lacY* gene, which has been cloned into a recombinant plasmid and sequenced. By combining overexpression with the use of a specific photoaffinity-labeled substrate for the permease and reconstitution of transport activity in artificial phospholipid vesicles (i.e., proteoliposomes), the permease was solubilized from the membrane, purified to homogeneity, and shown to catalyze all the transport reactions typical of the β-galactoside transport system *in vivo* with similar turnover numbers. Therefore, a single enzyme—the product of *lacY*—is solely responsible for all of the translocation phenomena catalyzed by the β-galactoside transport system. In addition, evidence has been presented that the permease is functional as a monomer.

Based on circular dichroic measurements indicating that purified permease is about 80 percent helical and on hydropathy analysis of the deduced amino acid sequence, a secondary structure was proposed. The model predicts that the protein has a short hydrophilic amino terminus, 12 hydrophobic domains in α-helical conformation that traverse the membrane in zigzag fashion connected by hydrophilic loops, and a 17-residue hydrophilic carboxyl-terminal tail. Spectroscopic, biochemical, and immunological data are consistent with the general features of the model and indicate that the amino and carboxyl termini are on the cytoplasmic surface of the membrane. Studies on an extensive series of lac permease–alkaline phosphatase chimeras have provided strong support for the topological predictions of the 12-helix model.

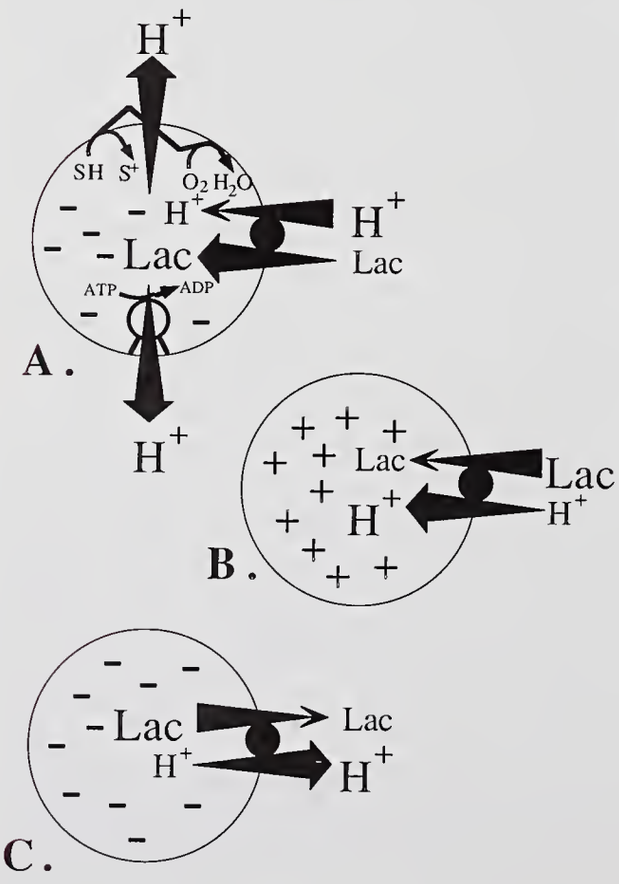
This report concentrates on current experiments that involve the use of site-directed mutagenesis to engineer lac permease so as to permit certain biochemical and biophysical approaches to structure-function relationships.

Recent studies provide definitive support for the argument that cysteinyl residues do not play a direct role in the lac permease mechanism. Thus, when site-directed mutagenesis is used to replace each of the eight cysteinyl residues simultaneously, the "C-less" permease catalyzes active lactose transport moderately well relative to wild-type permease (about 35 percent of the maximum velocity and 55 percent of the steady-state level of accumulation). Moreover, active lactose transport in right-side-out vesicles containing C-less permease is not inactivated by the alkylating agent, *N*-ethylmaleimide, in dramatic contrast to vesicles containing wild-type permease.

Although site-directed mutagenesis is useful for delineating amino acid residues that are important for lactose-H^+ symport and/or substrate binding and recognition, it has become apparent that high-resolution structure is required to begin to determine the role of these residues in the mechanism. Moreover, dynamic information at high resolution will also be required to solve the mechanism. In this respect, chemical labeling and spectroscopic approaches in which reactive cysteinyl residues are tagged with radioactive sulfhydryl reagents, electron paramagnetic labels, or fluorescent probes are potentially powerful means for examining static and dynamic aspects of protein structure-function relationships at high resolution. A principle difficulty with the general approach, however, is the complexity resulting from the presence of multiple cysteinyl residues in most proteins, eight in the case of lac permease. Thus, in addition to the important conclusion that cysteinyl residues do not play an important role in the mechanism of lac permease, the construction of a functional permease molecule devoid of cysteinyl residues provides the basis for an approach to the analysis of static and dynamic aspects of permease structure-function relationships.

By using the *lacY* gene encoding C-less permease, for instance, it is now possible to design mutants in which an individual amino acid residue in a putative hydrophilic or hydrophobic domain is replaced with a cysteinyl residue. This can then be reacted specifically with either permeant or impermeant sulfhydryl reagents in right-side-out or inside-out membrane vesicles, followed by solubilization and immunoprecipitation. In addition, single Cys mutants can be solubilized and purified, tagged with appropriate electron paramagnetic or fluorescent probes, then reconstituted and studied spectroscopically.

Finally, it should be possible to study proximity relationships between transmembrane domains by placing single cysteinyl residues in pairs of helical domains predicted to lie close to each other within the membrane. In these contexts, it is encouraging that more than 200 single Cys replacements have been constructed in the C-less permease and that the great majority of the mutants exhibit highly significant transport activity.



Lactose transport in Escherichia coli. *A: Uphill lactose (Lac) transport in response to H^+ generated either by respiration or ATP hydrolysis. B: Uphill H^+ transport in response to an inwardly directed lactose gradient. C: Uphill H^+ transport in response to an outwardly directed lactose gradient.*

From Kaback, H.R. 1989. Harvey Lect *83:77–105. Copyright © 1989 Alan R. Liss, Inc. Reprinted by permission of Wiley-Liss, a division of John Wiley and Sons, Inc.*

Control of the Immunoglobulin Heavy-Chain Gene

Thomas R. Kadesch, Ph.D.—Associate Investigator

Dr. Kadesch is also Associate Professor of Human Genetics at the University of Pennsylvania School of Medicine. He received his Ph.D. degree in biochemistry from the University of California, Berkeley, where he studied with Michael Chamberlin. His postdoctoral research was done with Paul Berg at the Stanford University School of Medicine.

EXPRESSION of immunoglobulin genes is limited to one cell type, namely B lymphocytes. Only B cells express the functions required for gene rearrangement, the process whereby mature immunoglobulin genes are formed from discrete gene segments; and only B lymphocytes possess the components necessary for immunoglobulin gene transcription, the process that creates an RNA copy of the rearranged genes.

Within these genes, there are at least two major transcriptional regulatory elements: the promoter, a DNA domain located close to the site where transcription begins, and the enhancer, another DNA domain that stimulates initiation from the promoter. The activities of each of these elements are restricted to the B cells and are controlled by proteins that bind them. We have focused our studies on the proteins binding the immunoglobulin heavy-chain (IgH) gene enhancer, in an attempt to understand how they elicit the enhancer's B cell–specific activity.

Contributions from a number of laboratories, including our own, led to a detailed map of the IgH enhancer. The map indicates the protein-binding locations and gives some information about the regulatory mechanism. The enhancer, while relatively small (200 base pairs), proves to be exceedingly complex. Interestingly, many of the perhaps nine or more distinct enhancer-binding proteins are found in multiple cell types, even those in which the enhancer is normally inactive. To explain this, it has been argued that some of the proteins may serve to stimulate the enhancer's activity, while others may repress it.

During the past few years, our efforts have been directed toward the isolation of cDNAs that encode these enhancer binders. Thus far, we have identified six. We are presently using segments of these genes (cDNAs) to manipulate and characterize their encoded proteins both structurally and functionally.

We have shown that two of the IgH enhancer-binding proteins, E2-5 and TFE3, are involved in a fascinating transcriptional regulatory scheme. In B cells the situation is relatively straightforward, as both proteins bind the enhancer and act in concert to stimulate its activity. In non–B cells the situation is more complicated. *In vivo* experiments suggest the presence of an additional protein, a repressor, that binds the enhancer and possibly precludes E2-5 action.

Binding of this putative repressor has two effects. First, the enhancer is less active as a result of the absence of bound E2-5 protein. Second, the repressor can attenuate at a distance the function of TFE3. Hence the presence of the repressor in non–B cells results in the shutdown of both E2-5– and TFE3-mediated activation. The effects of the repressor can be overcome in these cells by artificially overproducing the E2-5 protein. Presumably such overexpression is sufficient to displace the bound repressor.

A third gene segment isolated in the laboratory likely encodes the repressor. Although we have yet to prove this directly, the encoded protein, denoted Zeb, is structurally distinct from the aforementioned proteins that activate transcription and, moreover, falls into the family of so-called zinc finger proteins, many of which have been shown to repress transcription of other genes. The DNA sites to which Zeb binds in the test tube overlap, but are distinct from, those that the activator proteins bind, yet are identical to those that confer the repressing activity within cells. Ongoing experiments should further characterize Zeb's mode of action and possible roles in other gene systems.

E2-5 can be specifically repressed by yet another protein, a structural relative named Id (identified and initially characterized by Harold Weintraub [HHMI, Fred Hutchinson Cancer Research Center]). Unlike Zeb, Id binds directly to E2-5 (and related proteins), prevents its binding to DNA, and thus keeps it from stimulating transcription. Id expression was observed to fall off in several cell lines when they were induced to differentiate, suggesting that it may serve as a general antagonist of cellular differentiation by inhibiting DNA-binding proteins specifically required for differentiation.

The importance of E2-5 for IgH enhancer activity led to our asking whether Id plays a role dur-

ing B lymphocyte maturation. In collaboration with Stephen Desiderio (HHMI, Johns Hopkins University School of Medicine), we analyzed Id mRNA expression in a variety of cell lines representing different stages of B lymphoid–cell differentiation. Only two of these lines, representing the earliest stages of development, were found to express Id. We confirmed that these cells were unable to support the activity of the IgH enhancer, whereas later-stage cells could. Hence Id is restricted to early B lymphoid–cell progenitors, and its presence correlates inversely with IgH enhancer activity. Presumably Id serves to keep enhancer activity low in the B cell progenitors. Further differentiation of B cells would require a decrease in Id expression that, in turn, would allow the IgH enhancer to become active.

We have shown that Id also plays an important role in the development of another hematopoietic lineage, the myeloid lineage. In collaboration with Giovani Rovera (Wistar Institute), we demonstrated that Id is expressed in neutrophil precursors. When induced to differentiate in culture, the Id levels in these cells were observed to decrease, but only transiently. Cells bearing an artificial, constitutively expressed Id gene failed to differentiate. Hence an appropriate, transient shutoff of Id is required for neutrophil development.

An important corollary to these results is that myeloid cells must utilize proteins of the E2-5 family to promote their differentiation. We are presently attempting to identify and characterize these proteins.

Genetic Control of Hemoglobin Synthesis

Yuet Wai Kan, M.D., D.Sc.—Investigator

Dr. Kan is also Louis K. Diamond Professor of Hematology in the Departments of Laboratory Medicine and Medicine at the University of California, San Francisco. He received his M.D. and D.Sc. degrees from the University of Hong Kong Medical School. After internship and residency at Queen Mary Hospital, Hong Kong, he obtained postdoctoral training in hematology at Peter Bent Brigham Hospital, the Massachusetts Institute of Technology, Royal Victoria Hospital at McGill University, and the Children's Hospital, Boston. Dr. Kan has received numerous honors, including the Gairdner Foundation International Award and the Albert Lasker Clinical Medical Research Award. He is a fellow of the Royal Society (London) and a member of the National Academy of Sciences and of the Academia Sinica (Taiwan).

THE focus of our research is the molecular basis of genetic diseases affecting the hematopoietic or blood-forming cells. The two diseases we have studied in depth are sickle cell anemia and thalassemia. Both result from abnormal globin production and constitute important health problems in the Mediterranean region, Africa, the Middle East, and Asia. In the United States these disorders occur frequently among people of African, Italian, Greek, and Asian descent. We have defined the mutations that give rise to these defects and devised DNA analyses for their detection. In addition, we are studying the control of globin gene expression in red cell precursors and the signals that switch these genes from fetal to adult globin production.

Previously we demonstrated that the common genetic defect in α-thalassemia is deletion of the α-globin structural gene. We also defined some of the molecular lesions in β-thalassemia. These studies led to our ability to detect thalassemia by analysis of fetal DNA.

We initiated a new method of linkage analysis using restriction endonucleases to detect polymorphism in DNA sequences and applied it to tracing the evolution of the sickle and thalassemia mutations. Restriction enzyme site polymorphism is now an important tool for detecting many genetic disorders and for mapping the genetic loci of many diseases.

We developed a method for prenatal diagnosis of sickle cell anemia and thalassemia. Initially, fetal samples were required. With the advent of recombinant DNA technology, mutations in the human genome can be analyzed directly using DNA obtained by amniocentesis or chorionic villus biopsy, permitting early *in utero* diagnosis of these conditions.

Prenatal Diagnosis of Sickle Cell Anemia and Thalassemia

The polymerase chain reaction has made it possible to diagnose many point mutations rapidly. For sickle cell anemia, practical nonradioactive tests are now available. In the case of β-thalassemia, however, the need to diagnose multiple mutations makes these tests somewhat tedious, especially in the developing countries where the disease is common. Hence we are now devising a rapid approach using the reversed dot blot principle. Oligonucleotides corresponding to the mutations common in a given area are immobilized on a filter; the test DNA is amplified and hybridized to the oligonucleotides; and one of several nonradioactive methods is then used for detection. We believe this procedure will facilitate prenatal diagnosis of β-thalassemia on a broad scale.

Control of Globin Gene Expression

We are studying the factors that control the tissue- and development-specific expression of the human globin genes. Synthesis of the globin chains is precisely coordinated during development. The embryonic ϵ- and ζ-globins, which are synthesized in the early embryo, are replaced in the fetus by the α- and γ-globin chains. Prior to birth, the β-globin chain is known to become predominant over the γ-globin chain; but the factors that control the expression of the β-globin gene in the bone marrow cells and coordinate the expression of the globin genes during development have not yet been elucidated. We are now studying the DNA sequences and protein factors that govern the expression of these genes.

Recently DNA elements that are important for the control of the tissue-specific and development-specific expression of the globin genes have been revealed. This sequence is known as the locus control region, or LCR. In the β-globin gene, the LCR contains four hypersensitive sites, which contain consensus sequences that bind trans-acting factors. We have been studying the cis sequences, which are important for protein-DNA interactions in two ways.

First, we have determined by mutation analysis of these sequences, using both DNA-transfection and transgenic experiments, that important core

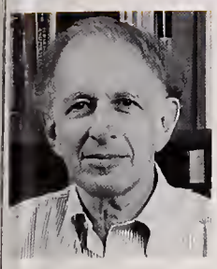
elements called NFE2-AP1–binding sites are essential for these enhancing activities. In addition, using *in vivo* footprint experiments, we have examined the consensus sequences that are important for *in vivo* interaction. We have studied several cell lines that express different amounts of ϵ-, γ-, and β-globin genes and have found that the consensus sequences are bound differently in these cell lines, depending on the type of β-globin gene expressed. We are now characterizing the factors responsible for this binding in order to understand the molecular basis of hemoglobin switching.

Red Cell Membrane Disorders

Many hereditary hemolytic anemias are accompanied by spherocytosis or elliptocytosis. The defects in these disorders lie in mutations of various cytoskeleton proteins, such as spectrin, ankyrin, and protein 4.1. We have studied families with hereditary elliptocytosis due to protein 4.1 deficiency and have defined three different defects due to gene rearrangements.

Protein 4.1 is found in all cell types and exists in multiple isoforms generated by alternate splicing in at least five different regions of the gene. Some isoforms are more abundant in nucleated cells, others in mature red cells. Although protein 4.1 maintains the integrity of the cytoskeleton in the red cell, its function in nucleated cells is not yet known. In a patient who has no protein 4.1 in the red cell, severe hemolytic anemia ensues, while other cellular functions are normal. Structural analysis showed that the protein 4.1 gene is rearranged at the region corresponding to the amino terminal of the erythroid protein 4.1. Otherwise it continues to produce functional protein 4.1 by an alternate splicing mechanism, which skips the mutated region. This example may point out an advantage of alternate splicing.

Cell Biological Studies of Memory



Eric R. Kandel, M.D.—Senior Investigator

Dr. Kandel is also University Professor of Physiology and Psychiatry at the Center for Neurobiology and Behavior of the Columbia University College of Physicians and Surgeons. He was born in Vienna, Austria. He graduated from Harvard College, having majored in history and literature, and received his medical degree from the New York University School of Medicine. He took postdoctoral training with Wade Marshall in the Laboratory of Neurophysiology at NIH and with Ladislav Tauc at the Institut Morey in Paris. Dr. Kandel was the founding director of the Center for Neurobiology and Behavior at Columbia. He is a member of the National Academy of Sciences and counts among his honors the Lasker Award, the Gairdner Award, and the National Medal of Science.

LEARNING is commonly divided into two major types: declarative and reflexive. Declarative learning refers to the acquisition of information about people, places, or things. Reflexive learning refers to the acquisition of motor skills and strategies. Our laboratory has been studying elementary forms of reflexive learning in the gill-withdrawal reflex of the marine snail *Aplysia*. In an attempt to compare these mechanisms with those underlying declarative forms of learning, we are also conducting a study of long-term potentiation in the mammalian hippocampus. The present discussion is limited to reflexive studies in *Aplysia*.

We have shown that the gill-withdrawal reflex can be modified by nonassociative and associative forms of reflexive learning, giving rise to both short- and long-term memory, whose duration is a function of the number of training trials. We have recently focused on sensitization, a nonassociative form of learning in response to a noxious stimulus. To analyze the relationship between the memory for short- and long-term sensitization, we have studied in particular detail one component of the neural circuit of this reflex: the connections between the siphon sensory neuron and the gill motor neurons. These connections can be studied both in the intact animal and in dissociated cell culture.

With sensitization the connections undergo an increase in synaptic effectiveness (facilitation), whose duration is a function of the number of behavioral reinforcing stimuli to the tail. Similarly, in culture, the duration of the facilitation is a function of the number of applications of serotonin (5-HT), a modulatory transmitter released by tail stimuli. A single tail stimulus or a single pulse of 5-HT produces short-term facilitation lasting minutes, whereas continuous application of 5-HT for 1.5 h or four or five pulses over a 1.5-h period elicit long-term facilitation lasting one or more days.

In both the behavioral reflex and the monosynaptic facilitation, there is a parallel requirement for protein and mRNA synthesis for long-term but not for short-term effectiveness. The short-term process reflects enhanced transmitter release from preexisting synaptic connections due to covalent modification of preexisting proteins. The long-term process requires new protein synthesis and leads to the growth of new synaptic connections.

Proteins and mRNAs necessary for long-term memory must either be induced during the brief (1.5-h) time window of training or be constitutively expressed and transiently accessible to covalent modification during this period. The experiments described below were designed to distinguish between these two possibilities.

In earlier experiments, Ari Barzilai, Tim Kennedy, David Sweatt, and I found that repeated pulses of 5-HT induced changes in the synthesis of specific proteins in the sensory neurons. Similar changes in protein synthesis can be produced by cAMP, a second messenger activated in the sensory neurons by 5-HT. These proteins could therefore reflect the transcription of cAMP-inducible genes. In mammals, genes induced by cAMP share a control element called the cAMP recognition element (CRE), which binds transcriptional activators called the CRE-binding proteins (CREBs). *Aplysia* neurons contain proteins homologous to mammalian CREBs.

These findings raise three questions that Bong-Kiun Kaang, Seth Grant, and I attempted to address: Can the facilitating transmitter 5-HT induce transcriptional activation of reporter genes in the sensory neurons that are driven by the CRE? Does this transcriptional activation correlate with the graded induction of long-term facilitation? and How does 5-HT activate transcription by CREB?

Because the CRE will also confer cAMP inducibility when placed upstream of reporter genes such as β-galactosidase, these gene constructs can serve as an assay system for transcriptional induction by cAMP. Kaang, Grant, and I therefore microinjected a CRE-*lacZ* reporter plasmid into

Aplysia sensory neurons and measured the levels of expression driven by the CRE in response to stimulation by 5-HT. We next exposed the neurons to five repeated (5-min) pulses of 5-HT, the protocol that produces long-term synaptic facilitation, and found a fourfold induction of the reporter gene. Similarly, cAMP induced expression by 3.4-fold. Thus repeated pulses or prolonged exposure to 5-HT stimulates transcription, and this stimulation can be simulated by elevating levels of cAMP. The increase in *lacZ* expression is mediated specifically through the CRE sequence. A plasmid (ΔCRE-*lacZ*) from which the CRE element is deleted shows no expression.

Does the stimulation of transcription mediated by 5-HT depend on the binding of transcription factors to the CRE sequence? To address this question, we co-injected the CRE reporter construct with either an oligonucleotide that encodes the wild-type CRE sequence and binds CREB, or with a mutant oligonucleotide that does not bind CREB-like factors. The wild-type oligonucleotide, which binds to and titrates out CREB, blocked the 5-HT–induced expression of the reporter construct, whereas the mutant oligonucleotide produced no inhibition.

These experiments suggest that the induction mediated by 5-HT requires positively acting CREB-like cellular factors that interact with the CRE, and provide direct evidence that 5-HT can stimulate transcription of genes containing the CRE sequences. This is consistent with results that Pramod Dash, Binyamin Hochner, and I obtained last year in physiological experiments, in which we microinjected the CRE into the nucleus of sensory neurons and selectively blocked the long-term increase in synaptic effectiveness without affecting short-term facilitation. Both these sets of results suggest that CREB-like transcriptional activators are required for the induction of long-term facilitation.

Does the activation of transcription by 5-HT have a clear threshold or is it graded? To assess this question, Kaang, Grant, and I were influenced by our earlier findings discussed above, that a single (5-min) pulse of 5-HT will only elicit short-term facilitation. To generate significant long-term facilitation, four or five pulses are required. Using the same pulse protocol, we found that a single pulse of 5-HT does not, in fact, stimulate expression of the reporter gene, whereas four to five pulses of 5-HT/IBMX produced a sevenfold stimulation. An intermediate number of pulses (two or three) gave an intermediate level (twofold) of stimulation. This correlation, between the number of 5-HT trials and the level of transcriptional induction and facilitation, suggests that the graded nature of long-term facilitation may reflect the graded nature of transcriptional induction.

Does activation by 5-HT require that CREB be phosphorylated? If so, must activation of CREB in the sensory neuron be mediated by protein kinase A (PKA)? To address these questions, Kaang, Grant, and I next microinjected two constructs into the sensory neurons: a reporter gene as well as a chimeric transactivation plasmid that expresses CREB. Following co-injection, exposure to 5-HT produced a 10-fold stimulation of transcription. By contrast, injection of either transactivation or reporter plasmid alone showed no expression. This transcriptional stimulation depends on phosphorylation of the PKA consensus site at Ser^{119} in CREB. Microinjecting a transactivation plasmid (pSA119) containing a mutation that converts Ser^{119} to Ala^{119} showed no stimulation with 5-HT.

Thus long-term facilitation induced by 5-HT leads to activation of CRE-inducible genes. The induction of these genes by the cAMP cascade is graded and can be initiated by PKA through phosphorylation of CREB. A conventional modulatory transmitter, 5-HT, can therefore select either a cytoplasmic or a genomic program of cellular action, depending on the number of presentations of 5-HT. By being able to activate a nuclear signal (through the phosphorylation of CREBs and its action on the CREs), modulatory transmitters can activate transcription and thereby take on the properties of growth factors. In the case of sensory neurons, transcriptional activation of CREB-related proteins seemed to represent one component of the switch, which extends the short-term cytoplasmic process for synaptic facilitation into the genomic changes characteristic of the long-term facilitatory process—one that includes the growth of new synaptic connections.

The T Cell Repertoire

John W. Kappler, Ph.D.—Investigator

Dr. Kappler is also a member of the Division of Basic Immunology of the Department of Medicine at the National Jewish Center for Immunology and Respiratory Medicine, Denver, and Professor of Microbiology and Immunology and of Medicine at the University of Colorado Health Sciences Center, Denver. He was educated at Lehigh University and received his Ph.D. degree in biochemistry at Brandeis with Gordon Sato. He did postdoctoral work at the University of California, San Diego, with Richard Dutton. After holding faculty positions at the University of Rochester, he moved to his present position at the National Jewish Center. He was awarded the Wellcome Foundation Prize by the Royal Society and is a member of the National Academy of Sciences.

As protection against invasion by foreign organisms, higher animals have evolved a complex collection of cells and chemicals broadly termed the immune system. Components of this system are able to recognize foreign material in the body and give rise to a series of events that cause the destruction or inactivation of the invader. Cells arising in the thymus, the T lymphocytes, are central to the efficient function of this system.

T cells bear receptors on their surfaces that are able to interact specifically with foreign material. Such interaction stimulates these cells to produce chemicals that allow other cells, and the T cells themselves, to respond to the invader. There are two kinds of T cells, bearing $\alpha\beta$ or $\gamma\delta$ receptors respectively. The $\alpha\beta$ receptors are made up of several segments—Vα, Jα, Vβ, Dβ, and Jβ—each of which can differ in structure from one T cell to another. This is possible because the DNA of higher mammals contains a number of alternate genes for each of these segments. As each T cell develops, it selects a different combination of these genes, and therefore eventually expresses $\alpha\beta$ receptors that are not exactly the same in structure as those of its fellows. It is these variations in $\alpha\beta$ receptor sequence that enable one T cell to recognize influenza virus, for example, and another poliovirus.

In order for T cells to recognize most foreign materials, they must bear exactly the right combination of variable segments—Vα, Jα, etc. The proper combinations are usually rare, so when an animal is confronted with an invading organism, the cells that can actually recognize the invader are few, probably about one in 100,000 or one in a million of all T cells. This fact does not hold true, however, for special types of foreign material called superantigens.

Superantigens bind to special cell-surface molecules—class II proteins of the major histocompatibility complex (MHC). They then interact with the Vβ portion of the T cell receptor, almost without regard to the composition of its other variable elements. Since there are only about 75 different sequences for Vβ in humans, any given superantigen will, theoretically, react with at least 1 percent of all T cells. In the mouse, with fewer different Vβ sequences, superantigens react with at least 5 percent of T cells. In fact, a particular superantigen can react in some cases with up to 30 percent of all T cells in either of these species.

The fact that superantigens can react with so many T cells causes them to have some important pathogenic properties. For example, massive stimulation of T cells by superantigens causes toxic shock in humans, and there is reason to believe that these antigens may be involved in certain autoimmune diseases.

Our laboratory has recently been studying the interaction among superantigens, the Vβ portions of T cell receptors, and the class II MHC proteins. A staphylococcal toxin, SEB, has been used as a model for these experiments. Amino acids have been identified in SEB that control the binding of this protein to Vβ or class II.

In many cases proteins that must bind to two different target molecules express their binding sites in different domains. That is, the different functions of the protein are separated spatially. Surprisingly, this does not seem to be the case for SEB. The binding sites of this protein for Vβ and class II appear to be interwoven, as though they must lie in close proximity.

The structural studies on SEB have led to the creation of a collection of mutant SEBs, some able to bind class II MHC but not Vβ, some with a more limited range of Vβ specificities than SEB itself, and some able to bind neither MHC nor the T cell receptor. These mutant SEBs are now being screened as vaccines. Mice preimmunized with the mutant molecules are no longer sensitive to the toxic effects of SEB given later.

Our laboratory and others have discovered a second class of superantigens, encoded by retroviruses that cause mammary tumors in the mouse. Although the genes coding for these superantigens have been known for some time, the struc-

ture and function of their protein products have only recently been determined. They turn out to be unusual proteins. Although they are bound to the surface membranes of cells, they differ from most surface proteins, in that the amino acids of their carboxyl terminal rather than amino terminal lie outside the cell. Only a small amino-terminal stretch of the protein lies inside. Experiments show that this inside portion is not essential for the superantigenic properties of the viral proteins, but that the proteins must be membrane bound in order to engage T cell receptors. The extreme carboxyl-terminal set of amino acids are among those that bind to T cell receptor Vβs.

In attempts to find out more about how these viral superantigens operate, monoclonal antibodies have been raised to various parts of the proteins. The antibodies reveal that only a few of these proteins are expressed on the cell surface—probably about 1,000 per cell. Despite this small number, the viral superantigens are extremely effective stimulators of T cell responses.

The Genetic Control of Morphogenesis

Thomas C. Kaufman, Ph.D.—Investigator

Dr. Kaufman is also Professor of Genetics in the Department of Biology at Indiana University, Bloomington, and Adjunct Professor of Medical Genetics in the Department of Medical Genetics at Indiana University Medical Center. He received his M.A. and Ph.D. degrees from the University of Texas, Austin, and did his postdoctoral research at the University of British Columbia in Vancouver.

THE long-term goal of our laboratory is to contribute to an understanding of the genetic basis of the developmental program of higher organisms. The organism chosen for our studies is the fruit fly *Drosophila melanogaster*, and our principal focus is a set of genes called homeotic, which play a crucial role in development.

The homeotic genes were first identified by virtue of the striking phenotypes observed when flies carried mutations at these loci. Specifically these homeotic lesions cause one portion of the animal to be transformed into an identity normally found in another region. Thus mutations at the *Antennapedia* locus cause a transformation of the antennae of the adult fly into a leg, and lesions in the *proboscipedia* gene result in the development of legs in place of the adult mouthparts. Both genes are members of a cluster of five homeotic genes called the *Antennapedia* complex (ANT-C), which is found in a restricted domain at the base of the right arm of chromosome 3.

The aggregate results of genetic, developmental, and molecular analyses of the ANT-C have revealed that the role of the resident loci is best viewed as a series of developmental switches for either/or decisions of cellular fate in the embryonic and larval stages of the organism. Furthermore, DNA sequence analysis of the homeotic loci reveals that they encode proteins containing a motif, dubbed the homeodomain, that endows the proteins with DNA-binding ability. Indeed, the homeotic proteins are found complexed with the nuclear DNA of the cells in which they are expressed. Thus it appears that the switch activity of the loci is reflected in their functioning as regulators of specific target genes. Not entirely clear at this point is how each homeotic locus is restricted to its own unique pattern of expression and which sets of genes are the targets of the switches.

In order to investigate these two unknowns, we have concentrated our efforts on three of the resident members of the ANT-C: *Sex combs reduced* (*Scr*), *proboscipedia* (*pb*), and *labial* (*lab*). Each gene was chosen for the unique properties it displayed during initial characterization. For example, *Scr* is the only homeotic gene expressed at the juncture between the head and trunk of the developing animal, and there were indications that the regulatory hierarchy of genes expressed in these two domains is different. Additionally, a genetic analysis revealed *pb* and *lab* to be small by homeotic standards. This meant that for the two genes, a complete dissection of the regulatory elements of each locus was feasible.

The *Sex combs reduced* Gene

Our prior genetic and molecular analyses of the *Scr* locus had shown that DNA sequences 50 kilobases (kb) distal to the point at which the RNA product is initiated were necessary for normal gene expression. Using "enhancer sniffers"—constructs capable of detecting DNA fragments that have the ability to regulate gene expression—we surveyed the entire *Scr* locus for such elements. To date we have found at least five fragments that specify the accumulation of gene product in the posterior head and anterior thorax. These are scattered over a 30-kb interval; consistent with our earlier genetic results, the regions of DNA that control head and trunk expression appear to be physically separate.

Genetic analysis has also shown that *Scr* is subject to the regulatory effects of "transvection." Normally the gene is only expressed in the first, or most anterior, thoracic segment; however, certain mutants in the gene allow its product to accumulate more posteriorly, in the second and third thoracic segments. We have shown that the normal restriction of pattern results from negative regulation and requires that the two copies of the gene in the cells of the posterior thorax be paired with each other. If this pairing is disrupted, the negative effect is removed and abnormal ectopic expression occurs.

Using sniffer constructs similar to those above, we have identified three DNA fragments that appear to be associated with the transvection effects at the *Scr* locus. Two of these three elements are located approximately 40 and 10 kb upstream of the gene, while the third is located within an in-

tron about 15 kb downstream of the transcription start site. These analyses of *Scr* are supported by a grant from the National Institutes of Health.

The location of these three elements relative to the identified enhancer elements is intriguing. Two flank the enhancers, while the third is located in the midst of the regulatory elements. Our current hypothesis is that the transvection elements serve to define a chromatin domain controlling an on/off state for the locus and that the resident enhancers can only have their effects in the on state. We are currently extending our analysis to define more precisely the boundaries and sequences associated with the above-identified fragments. We also hope to identify further the cellular factors that act through these elements.

The *labial* Gene

Using enhancer sniffers, we have identified a majority of the sequences needed for the normal expression of the *lab* gene. These are found upstream, either within 3.6 kb of the transcription start site or within the major 14-kb intron. Using a minigene construct that contains all of the upstream sequences but lacks the intronic elements, we have succeeded in rescuing the lethality and morphological anomalies associated with deletion of the *lab* gene in the embryo. Thus it would appear that the regulatory elements found within the intron are not necessary to *lab*'s normal embryonic functions.

Although the *lab* minigene rescues the embryonic defects associated with *lab*-deficient genotypes, there is no apparent rescue of adults. We have shown that this failure does not result from the minigene's inability to be expressed in adult tissues, but rather because its protein product is inappropriately and ectopically accumulated. This abnormal expression pattern only takes place in *lab* mutant animals and is not detected in normal minigene-bearing hosts.

We have found that the normal *lab* gene actually encodes two polypeptides, one of which is six amino acids longer than the other. The minigene is only capable of directing the synthesis of the shorter protein. It would appear, therefore, that animals capable of making the long form show correct adult expression, while short-form animals cannot, and that normal adult expression involves autogenous negative regulation.

Our earlier studies revealed that *lab* expression in the embryo also involved autogenous regulation. In that case, however, the effect was positive. That is, *lab* protein served to keep the *lab* gene turned on, not to turn it off. Moreover, the short-form protein was capable of performing this function.

We have now demonstrated that the sequences at the *lab* locus necessary for the positive and negative loops are, like the required proteins, different. The positive sequences are upstream of the transcription initiation site, while the negative targets are in the transcribed portion of the gene in the first exon. We are currently investigating the possibility that the two loops are affected by the direct interaction of the *lab* protein isoforms with the two alternate target sites in the *lab* gene itself.

The *proboscipedia* Gene

As in the case of *lab*, we have used a minigene derived from the *pb* locus and have obtained full rescue of the adult homeotic mutant phenotype. Moreover, fragments extracted from one of the introns and placed upstream of the *pb* promoter function to direct a normal spatiotemporal pattern of expression. Thus sequences both up- and downstream of the transcription start site are required for normal *pb* expression.

However, tests of the "enhancer" sequences on a heterologous promoter element in the sniffer constructs demonstrate no activity. On the other hand, minigene constructs in which the enhancer elements are deleted can be directed by novel "enhancer" elements derived from other loci. Therefore it appears that there are specific promoter/enhancer interactions at this locus and that the specificity appears to lie primarily with the "enhancer" elements.

We have also found a second regulatory element at the *pb* locus upstream of the transcription start site. However, this element, unlike the positive enhancer above, appears to be negative and serves to prevent ectopic expression of the gene. Moreover, like the *Scr* locus, it seems to have a pairing-sensitive component. We are now beginning a dissection of this system, which offers an opportunity to make direct comparisons between two similar regulatory schemes in the ANT-C. Our work on the *proboscipedia* gene is supported by a grant from the National Institutes of Health.

Protein Folding and Macromolecular Recognition

Peter S. Kim, Ph.D.—Assistant Investigator

Dr. Kim is also Member of the Whitehead Institute for Biomedical Research, Associate Professor of Biology at the Massachusetts Institute of Technology, and Assistant Molecular Biologist at the Massachusetts General Hospital, Boston. His undergraduate degree in chemistry was obtained at Cornell University, where he studied with George Hess. After receiving the Ph.D. degree in biochemistry from Stanford University, where he studied with Robert Baldwin, Dr. Kim moved to the Whitehead Institute for Biomedical Research as a Whitehead Fellow.

INFORMATION transfer in biology generally proceeds from DNA to RNA (transcription) and then from RNA to protein (translation). The linear, unfolded protein chains made during translation must fold into a three-dimensional shape to be functional. Although the basic mechanisms of transcription and translation are understood, at least in outline, the transfer of information from one to three dimensions—i.e., protein folding—remains a major unsolved problem in molecular biology. To understand protein folding is a prime objective of this laboratory.

A second effort is aimed at understanding the principles of macromolecular recognition: specific protein-protein interactions and interactions between protein molecules and DNA. These interactions are central to much of molecular physiology and developmental biology. We have focused on a structural motif called the leucine zipper, which occurs in several different DNA-binding proteins, including the products of some nuclear oncogenes.

A third and new effort is the *de novo* design of peptides and proteins.

Protein Folding

Much of our work in this area is centered on bovine pancreatic trypsin inhibitor (BPTI), arguably the protein most thoroughly characterized in biophysical terms. It is difficult to determine the structures of protein-folding intermediates, because protein folding is a cooperative process. Trapped disulfide-bonded intermediates, such as those identified in the early folding steps of BPTI, are often rather insoluble; this hinders detailed structural characterization by nuclear magnetic resonance (NMR). We have developed a peptide model approach that circumvents the problem of cooperativity and improves solubility, so that the structures contained within protein-folding intermediates can be characterized in detail.

Peptide models that simulate two crucial early intermediates in the folding of BPTI have been designed and synthesized chemically. By using two-dimensional NMR, we find that the structures contained within these peptide models are remarkably native-like, corresponding to subdomains of BPTI. These results suggest that a large part of the protein-folding problem can be reduced to identifying and understanding subdomains of native proteins.

Earlier work by others, however, concluded that there are well-populated, nonnative states in the oxidative folding of BPTI. This conclusion complicates efforts to understand protein folding. Recently we reexamined the spectrum and population of intermediates present during the folding of BPTI, taking advantage of improvements that have been made in separation technologies in the years since the original BPTI-folding experiments. In contrast to earlier studies, we find that all of the well-populated intermediates in the folding of BPTI contain only native disulfide bonds and that the essential features of the BPTI-folding reaction are determined in large part by native structure. These results emphasize the importance of native protein structure for understanding protein folding.

A recombinant model for an early predominant intermediate, containing a single disulfide bond between residues 5 and 55, has been made by replacing the cysteines not involved in the disulfide bond with alanine. Remarkably, this model folds essentially into the same conformation as native BPTI, as judged by two-dimensional NMR, and it inhibits trypsin. These findings provide an explanation for the properties of this intermediate in the folding of BPTI and demonstrate that the native fold of BPTI can be obtained without the assistance of nonnative disulfide species. The recombinant model also provides an attractive model system for studies of protein folding. This work is supported by a grant from the National Institutes of Health.

Other efforts are directed at evaluating electrostatic fields at the ends of α-helices and developing a model system to evaluate the β-sheet propensities of different amino acid residues.

Macromolecular Recognition

In this area, we have focused on the leucine zipper class of DNA-binding transcriptional acti-

vator proteins, originally identified by Steven McKnight and his co-workers (HHMI, Carnegie Institution). The leucine zipper regions of these proteins are important for homodimer or specific heterodimer formation.

Our approach in this work is to use "protein dissection." GCN4, a homodimeric transcription factor, serves as a prototype protein. A synthetic peptide corresponding to the 33-residue leucine zipper region folds as a parallel pair of helices. This led us to propose that leucine zippers are actually short coiled coils. X-ray crystallographic studies of this peptide (with Tom Alber's group, University of California, Berkeley) confirm that the leucine zipper of GCN4 is a coiled coil and provide the first high-resolution structure of a two-stranded parallel coiled coil. The effects of amino acid replacements on the stability, structure, and dynamics of this leucine zipper are being investigated.

Proper biological function requires that recognition between many different macromolecules in the cell occurs with exquisite specificity. We have found that the isolated leucine zipper regions from the nuclear oncogene products Fos and Jun are sufficient to mediate specific heterodimer formation. This provides a very simple model system for studying the specificity of protein-protein interactions: two helices that prefer to interact with each other rather than with themselves. By making hybrid leucine zipper peptides, we found that eight amino acid residues from each of the leucine zipper sequences are sufficient to mediate specific heterodimer formation. The predominant mechanism for specificity in this system was found to be electrostatic in nature.

A region of GCN4 rich in basic amino acid residues, immediately adjacent to the leucine zipper, is involved in DNA recognition. We find that this basic region by itself, when dimerized via a flexible disulfide linker in place of the leucine zipper, is also capable of sequence-specific DNA binding. In addition to simplifying structural analysis of this new DNA-binding motif, the finding provides a new strategy for the design of DNA-binding peptides. This work was supported by a grant from the National Institutes of Health.

Peptide and Protein Design

Knowledge of the rules involved in protein folding and macromolecular recognition can be tested by trying to design *de novo* amino acid sequences that fold into specific conformations and/or that interact in a predetermined manner with other molecules. Toward this end, we have begun to analyze the structural hierarchies in natural proteins and to design simple "building blocks" of peptide structure.

RNA Viral Genetics

Karla A. Kirkegaard, Ph.D.—Assistant Investigator

Dr. Kirkegaard is also Assistant Professor of Molecular, Cellular, and Developmental Biology at the University of Colorado at Boulder and Adjunct Assistant Professor of Cellular and Structural Biology at the University of Colorado Health Sciences Center, Denver. After receiving a B.S. degree in genetics from the University of California, Berkeley, she developed her doctoral thesis in the Department of Biochemistry and Molecular Biology at Harvard University with James Wang. Her postdoctoral work in virology was in association with David Baltimore at the Massachusetts Institute of Technology and the Whitehead Institute.

FOR numerous viruses and other subcellular parasites, RNA rather than DNA is the molecule used for storage and transmission of genetic information. We are interested in the consequences for a virus of having an RNA genome and in any mechanistic similarities or differences in genetic processes between RNA and DNA organisms. In addition, we are exploring the mechanisms of RNA packaging, replication, and recombination in the genome of poliovirus and other RNA viruses. We are also interested in the interactions between viruses and their host cells, especially in the realm of RNA-protein biochemistry.

Many of our genetic studies utilize poliovirus, a small icosahedral virus with an RNA genome of only 7,500 nucleotides. We have shown, for example, that RNA recombination occurs among poliovirus genomes with sufficient frequency that 1 out of every 25 is a recombinant. In contrast to the breaking and joining of molecules that leads to DNA recombination, recombination of RNA occurs during its synthesis. Genetic rearrangement results from the switching of parental templates by the viral RNA polymerase.

Thale Jarvis, a postdoctoral fellow in our laboratory, has developed a sensitive quantitative assay for RNA recombination using the polymerase chain reaction (PCR). Recombinants can be detected at a frequency as low as 1 in 10^6 parental genomes. Using this assay, we have been able to learn a great deal about the process of RNA recombination in poliovirus-infected cells. For example, we discovered that recombination frequency increases exponentially throughout the course of a single infectious cycle. It was surprising that the extensive cytopathic changes during the course of poliovirus infection do not inhibit the access of the parental RNA templates to each other. This work was also supported by a grant from the National Institutes of Health.

Taken together with other data we have obtained, our results suggest that the process by which recombinant RNA genomes are generated may be quite simple, and possibly universal among RNA viruses. We are using the PCR assay to test this idea by looking for RNA recombination among other RNA viruses that are less amenable to genetic analysis than poliovirus. We hope to gain a better understanding of the prevalence and mechanism of genetic recombination among RNA genomes, a process that is certainly responsible for much of the variability and rapid evolution of RNA viruses.

Using x-ray crystallography, James Hogle at Scripps Clinic has determined the three-dimensional structure of the poliovirion. However, an appreciation of functional interactions between the viral RNA and the virion proteins calls for the application of genetics as well as structural biochemistry. We do not know, for example, exactly which subviral protein particles package the poliovirus RNA into the final virion structure, nor do we know the structural requirements of the participants in the packaging reactions. Is the viral RNA threaded into an intact, preformed icosahedral capsid, or do smaller parts of the capsid condense around the viral RNA to form the final icosahedral structure?

We have constructed several mutants in the poliovirus RNA genome and have characterized them in great detail. Two of these mutants have pointed out a region of the viral capsid, quite internal to the virion, that is involved both in RNA packaging and RNA uncoating. We are examining the RNA-binding properties of subviral particles from both mutant and wild-type poliovirus-infected cells in hopes of identifying which subviral particles bind to RNA during intracellular assembly.

To investigate the RNA uncoating defects during cell entry of these mutant viruses, we have devised an assay to detect interactions between poliovirions and their cellular receptor directly. The laboratory of Vincent Racaniello (Columbia University) has identified the poliovirus receptor as a cellular adhesion protein of the immunoglobulin superfamily. Using antibodies prepared against peptides from the receptor's amino and carboxyl termini, we can analyze proteolytic digestion patterns of the receptor in the presence

and absence of bound wild-type and mutant poliovirions. In this way we are exploring the specific protein-protein contacts made between the viruses and the receptor during binding and the subsequent conformational changes during entry of the virus into the cell.

We have extended the study of RNA genetics to the interactions between yeast cells and the small double-stranded RNA genome of L-A, a cytoplasmic virus-like particle. Yeast cells, unlike the primate cells in which poliovirus and other RNA viruses of medical interest are propagated, are amenable to elegant genetic analysis, making it possible to identify quite quickly the cellular molecules that are involved in any given process. The virus-like particle L-A has been shown by the laboratory of Reed Wickner (National Institutes of Health) to replicate in the cytoplasm of yeast cells in large numbers. Genetic analysis of L-A, however, will depend on our ability to make defined mutations in its RNA genome.

In the past year, we have developed the technical capabilities to pursue this goal. First, we had to develop the technology to introduce RNA directly into yeast cells, which we accomplished by using RNA molecules encoding the luciferase protein of fireflies. Second, we needed an assay for the successful introduction of the L-A genome into yeast. Since the particle is so ubiquitous, no phenotype for its presence or absence in a yeast cell had been reported. We found that, under somewhat unusual laboratory conditions, certain yeast strains *require* the L-A genome for growth! This suggests the possibility of mutualistic relationships between host cells and virus-like RNA genomes, a possibility we are also pursuing in other systems.

Most importantly for this work, we now have a phenotype for the presence of the L-A RNA genome in yeast cells. We have recently been able to initiate L-A "infections" in yeast by introducing RNA made *in vitro*. We are now in a position to mutate this RNA genome, to study any defect in the L-A replication cycle caused by the mutations, and to observe the effects of yeast mutants on the replication cycle of mutant and wild-type RNA.

Adrenergic Receptor Structure and Function

Brian K. Kobilka, M.D.—Assistant Investigator

Dr. Kobilka is also Assistant Professor of Medicine, Cardiology, and Molecular and Cellular Physiology at the Stanford University Medical Center. He received his undergraduate degree in biology and chemistry from the University of Minnesota, Duluth, and his M.D. degree from Yale University. After his residency in internal medicine at Barnes Hospital, St. Louis, he joined the laboratory of Robert Lefkowitz as a research fellow in cardiology at Duke University. Four years later he was appointed Assistant Professor in the Department of Medicine at Duke University, and the following year he assumed his present position at Stanford.

THE autonomic nervous system serves as the master control center for the cardiovascular system. It monitors the effectiveness of the latter system in providing nutrients and oxygen to the rest of the body and appropriately adjusts the heart rate, blood pressure, and blood flow. These adjustments are made via nerves that serve the heart, blood vessels, and kidneys.

Adrenergic receptors form the interface between these nerves (of the sympathetic subsystem) and the organs they innervate. Catecholamines released from sympathetic nerve terminals bind to adrenergic receptors on the surface of target cells, and the activated receptors modify the function of these cells.

When a catecholamine occupies its binding site, the receptor activates a GTP-binding protein (G protein) inside the cell. The activated G protein may then modulate the activity of a cellular enzyme or ion channel. The genes (or corresponding cDNAs) for nine types of adrenergic receptors have been cloned. There are three types of α_1-adrenergic receptors, three types of α_2-adrenergic receptors, and three types of β receptors. All of these receptors are structurally similar, having seven hydrophobic domains that are thought to be membrane spanning. These features are shared by other receptors that activate G proteins.

α_2-Adrenergic Receptor Subtype Diversity

The role played by the α_1-, α_2-, and β-adrenergic receptors in the function of the sympathetic nervous system has been extensively studied. α_2-Adrenergic receptors are involved in blood pressure control and in directing blood flow to specific tissues. However, the physiological role of each of the three α_2 receptor types is not known. This is in part due to the lack of highly selective drugs that can activate or inhibit each type of α_2-adrenergic receptor. We are attempting to identify distinctive functional and physiological properties for each of the different α_2 receptors. These studies may provide incentive for the development of more highly selective α_2-adrenergic receptor drugs for the treatment of hypertension and vascular disease.

Adrenergic Receptor Structure

A major focus in my laboratory is to learn more about the three-dimensional structure of adrenergic receptors and to determine how they transmit signals across the cell membrane's lipid bilayer. We are taking several approaches to study the receptor structure. Over the past year our mutagenesis studies have identified a specific amino acid in the β_2-adrenergic receptor that forms part of the binding site for a large class of β receptor–blocking drugs. When this amino acid is placed in an α_2 receptor, it confers the ability to bind to these β receptor drugs. We have also used mutagenesis studies to identify intramolecular interactions within β_2- and α_2-adrenergic receptors. This information is useful in developing models for the three-dimensional structure of the receptor.

A long-range goal is to study the three-dimensional structure of the β_2-adrenergic receptor, using techniques that provide high-resolution pictures of this protein or detect changes in the structure of the receptor during signal transduction. These techniques require large quantities of pure, functional receptor protein. Using recombinant DNA techniques, we have made several modifications in the structure of the β_2-adrenergic receptor that lead to increased expression in cultured cells. Our goal during the next year will be to refine the purification procedure to optimize the yield and minimize the cost of large-scale production of receptor protein.

Receptor Biosynthesis

The primary amino acid sequence of a receptor contains all the essential information needed for the receptor's proper folding, post-translational processing, and cellular targeting. Understanding the process by which receptors are folded and processed should provide insights into receptor

structure and may identify factors that will enhance the production of functional protein.

Using a cell-free expression system developed in my laboratory, we have characterized the insertion of specific membrane-spanning domains into the lipid bilayer. The proper insertion of the first membrane-spanning domain is relatively inefficient. Using recombinant DNA techniques, we have identified a structural modification that enhances the integration of this domain and increases the amount of functional β_2 receptor protein expressed in cultured cells.

We have previously shown that β_2 receptor is nonfunctional immediately after translation and translocation into the endoplasmic reticulum. Our studies indicate that additional cytosolic and membrane factors are required to process the receptor to a functional form. We are attempting to identify structural differences between a functional receptor and newly synthesized receptors that have not undergone the post-translational processing necessary to produce functional protein. We hope to identify the requisite cytosolic and membrane factors.

Cellular Biology of β_2-Adrenergic Receptors

Following prolonged exposure to catecholamines, the β_2-adrenergic receptor becomes desensitized and is less efficient in activating adenylyl cyclase. Several mechanisms contribute to the process of desensitization, including receptor phosphorylation and the removal of receptors from the plasma membrane. Over the past year we have used immunocytochemical techniques to study agonist-mediated internalization of β_2-adrenergic receptors. These studies confirm that these receptors are rapidly internalized into intracellular vesicles called early endosomes. We plan to continue our efforts to determine the mechanism by which this occurs and to identify the cellular proteins that mediate this process. Furthermore, we hope to gain a better understanding of the relative importance of agonist-activated internalization in the physiology of β_2-adrenergic receptors.

Molecular Genetics of Lymphocyte Development and Neoplasia

Stanley J. Korsmeyer, M.D.—Associate Investigator

Dr. Korsmeyer is also Professor of Medicine and Molecular Microbiology at Washington University School of Medicine, St. Louis. He received his B.S. degree in biology from the University of Illinois, Urbana, and his M.D. degree from the University of Illinois, Chicago. He did his internship and residency in internal medicine at the University of California, San Francisco, and his postdoctoral research with Thomas Waldmann and Philip Leder at NIH, where he became a Senior Investigator at the National Cancer Institute. His honors include membership in the American Society for Clinical Investigation.

GENES that encode receptors for foreign antigens have provided our most pivotal insights into early lymphocyte development and lymphoid malignancies. The genes for immunoglobulin (Ig), or antibodies, and for the T cell receptor (TCR) encode the antigen receptors for B cells and T cells, respectively. During early lymphocyte development, recombination at the DNA level assembles these genes to create a wide repertoire of receptor specificities.

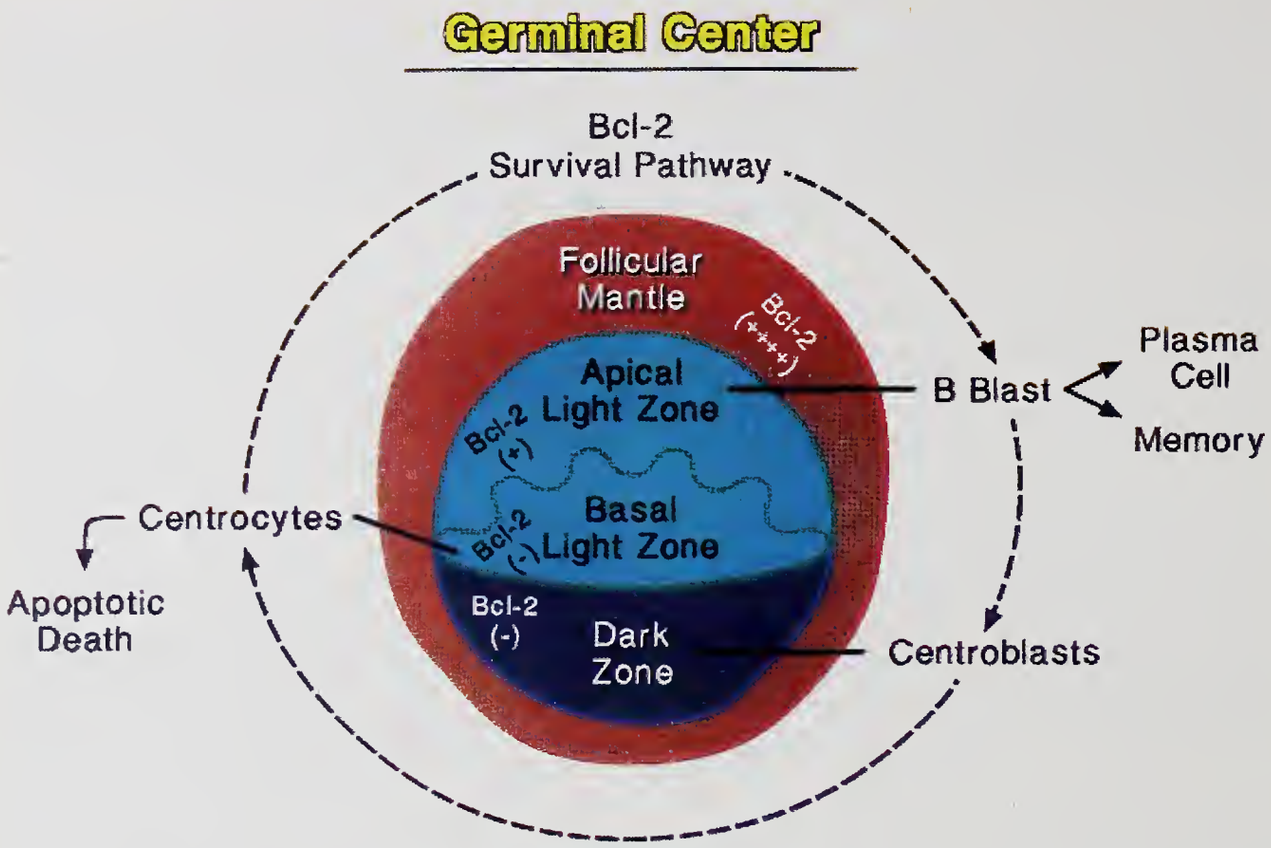
Much of what we know about these genes has been gleaned from studies of lymphoid tumors. These malignancies are clonal expansions of a single cell and provide multiple identical copies of these genetic events. Provocatively, the characteristic interchromosomal translocations that typify B cell malignancies break at the Ig genes, while those of T cell tumors often occur at the chromosomal home of the TCR genes. We have exploited this geography to clone the DNA at these illegitimate interchromosomal junctures. This serves as a bridge from the antigen receptor loci to the other chromosomal partner, which has often introduced a new cancer-promoting gene.

As a prototype, we have cloned the juncture between chromosomes 14 and 18 that is present in the most frequent form of human lymphoma, follicular-type B cell lymphoma. This translocation occurs early in the development of a B cell and introduces a newly discovered gene, *Bcl-2*, into the Ig locus. A hybrid *Bcl-2*–Ig fusion gene is created, resulting in the overproduction of *Bcl-2*. Transgenic mice were created that possess a copy of the abnormal *Bcl-2*–Ig fusion gene in their genetic material. They progressed from an indolent expansion of resting B cells to high-grade life-threatening lymphomas, recapitulating the natural course of the human disease and proving that this translocation causes malignancy. The *Bcl-2* protein is unique among proto-oncogenes by being located in mitochondria. Moreover, it has a novel function in that it blocks the programmed death of cells independent of promoting their growth. When deregulated, *Bcl-2* extends the survival of cells normally destined to die. *Bcl-2* constitutes the first member of a new oncogene category, regulators of cell death.

In a parallel set of experiments, unanticipated rearrangements into the δ TCR locus on chromosome 14 have identified the interchromosomal translocation sites that typify early T cell acute lymphoblastic leukemias. Two of these new genes, *Ttg-1* from chromosome 11 and *Hox-11* from chromosome 10, are not normally expressed in T cells. Instead their function is diverted from their normal sites to T cells. Transgenic mice reveal that redirecting these regulatory genes into T cells causes malignancy.

The majority of chromosomal defects, however, have no candidate gene at either side of the juncture. The responsible gene lies a considerable distance from known genes. One promising approach utilizes portions of chromosomes from yeast organisms to obtain and propagate long stretches of human DNA. This makes it possible to clone and map entire segments of human chromosomes, precisely linking known genes, and generating probes to search for new disease loci. This approach has pinpointed the chromosomal breakpoint and identified a candidate oncogene for an extremely aggressive leukemia of childhood.

This group of studies aims to improve our understanding of the genetic pathways of early T and B cell development, as well as the aberrancies that result in malignancy.



Distribution of the Bcl-2 protein that blocks cell death within the germinal center of the immune system. Bcl-2 is confined to survival zones. B cells in the follicular mantle are longer lived and recirculate. The growing centroblasts in the dark zone and the dying centrocytes in the basal light zone lack the protein. It returns in the apical light zone, where long-lived memory cells emerge to permit recall responses.

Research of Stanley Korsmeyer.

Molecular Genetics of Neuromuscular Disease

Louis M. Kunkel, Ph.D.—Investigator

Dr. Kunkel is also Professor of Pediatrics and of Genetics at Harvard Medical School. He received his B.A. degree from Gettysburg College and his Ph.D. degree in biology from the Johns Hopkins University. He took postdoctoral training with Brian McCarthy at the University of California, San Francisco, and with Samuel Latt at the Children's Hospital, Boston. He held appointments at Children's Hospital/Harvard Medical School before joining HHMI. His honors include the Gairdner Award and election to the National Academy of Sciences.

We continue to study the underlying cause of Duchenne and Becker muscular dystrophies. Our identification of the gene led rapidly to our description of the encoded protein dystrophin, which was found to be an unknown member of a family of proteins that includes the spectrins and α-actinins. Located at the inner face of the plasma membrane of myofibers, dystrophin is thought to confer strength to the membrane during muscle contraction and relaxation. Absence or abnormality of dystrophin at this location causes the myofiber degeneration of Duchenne and Becker dystrophy. Our work has led to improved diagnosis of the diseases and to testable ideas on therapeutic intervention.

Over the past year the laboratory has concentrated on identifying new members of the dystrophin family of proteins, in the hope that these can play a role in mitigating disease caused by dystrophin deficiency. These dystrophin-related proteins might themselves be involved in other neuromuscular diseases. By antigenic cross-reactivity, we have cloned the human microtubule-associated protein 1B (MAP-1B) and have mapped this locus in close proximity to mutations on chromosome 5 that cause a motor neuron degenerative disease, spinal muscular atrophy (SMA).

We have sequenced this human gene and have begun its direct mutational analysis in SMA patients. We have also identified at the locus two separate (CA)n repeat polymorphisms that have been shown, in collaboration with Conrad Gilliam, to be linked tightly to SMA mutations. In this genetic analysis, a few rare recombinations have been detected between SMA and MAP-1B, implying that MAP-1B may be very close, but not involved in the disease. As a result of this observation, we have cloned large segments of human DNA from the region of chromosome 5 as yeast artificial chromosomes (YACs) and have begun a search for other genes near MAP-1B, any of which might be the SMA gene.

By reduced stringency hybridization with dystrophin cDNA clones, a chromosome 6–encoded protein was identified by Kay Davies in Oxford. This dystrophin-related protein (DRP) is highly similar to dystrophin, and antibodies against it developed in our laboratory have revealed that the two proteins are almost identical in size. Immunolocalization of DRP has shown that it colocalizes with dystrophin in a developing myofiber. In a mature fiber, however, it is located only at the neuromuscular and myotendinous junctions of muscle. We are currently attempting to clone the entire coding sequence of DRP as cDNA. We believe that the sequence of this protein might reveal why it has such a specialized location in mature muscle and what role, if any, it might play in mitigating some of the effects of dystrophin absence in Duchenne dystrophy. We are also searching for other neuromuscular diseases that might be caused by DRP abnormalities.

One obvious set of dystrophin relatives for us to attempt to characterize were the α-actinins. A smooth muscle form had been cloned and localized to chromosome 14, but there was at least one other that had been identified from chicken muscle. Using a conserved motif of the known α-actinin amino acid sequence, we designed a degenerate set of PCR (polymerase chain reaction) primers that amplified a human muscle product with the sequence of an α-actinin. Using this small product as a hybridization probe, we screened a muscle cDNA library and obtained two classes of hybridizing human cDNA clones. Sequence analysis revealed that the clones contained two different α-actinins that were unique for humans. They are both muscle specific and are encoded from chromosomes 1 and 11, respectively. We are now developing antisera specific for each of the muscle α-actinin gene products and attempting to identify disease phenotypes that might involve α-actinin.

Abnormalities of dystrophin are easily detected at the protein level, and nearly 70 percent of the mutations that cause them have been shown to be deletions or duplications of some part of this extremely large locus. We have designed primers from dystrophin's nucleotide sequence to allow PCR amplification of specific regions of dystrophin's transcript. In analyzing these PCR prod-

ucts from patients who do not have a detectable deletion, we have found at least one mutation that involves splicing of dystrophin's primary transcript.

PCR amplification yielded a product of abnormal size that lacked exon 5. Sequence analysis of the patient's DNA from the region of this exon revealed an identical sequence to that of normals. We are currently cloning other surrounding DNA to identify the exact point where a mutation might effect the splicing of exon 5. We believe that mutations in splicing signals may lead to a better understanding of how this huge locus is processed into a mature mRNA transcript.

During our screening of different dystrophin antibodies for finding cross-reactive proteins, our carboxyl domain–directed antibodies detected a small protein (approximately 70 kDa). At the same time, David Yaffe's group in Israel reported observing a short dystrophin transcript in nonmuscle tissues. Our laboratory cloned this transcript from our human brain cDNA library, and sequencing of the cDNA clones revealed a divergence from expected dystrophin sequence at approximately exon 62. We are currently preparing antibodies against novel peptide sequences to determine if the shorter protein detected in brain tissue with antidystrophin antibodies is indeed a product of this novel transcript.

In addition to the published shorter transcript, we have also detected a slightly larger protein (approximately 90 kDa) in sciatic nerve. cDNA clones have been prepared and shown to diverge from expected dystrophin sequence at exon 54. Similar studies are under way for this novel transcript to determine the role these smaller proteins might play in the phenotype of Duchenne dystrophy.

Our aim for the future year is to build on work already in progress. We will continue to identify dystrophin-related proteins and attempt to determine their normal function in muscle. We will determine what role, if any, they play in other neuromuscular diseases and whether they mitigate the effects of abnormal dystrophin. We will continue our analysis of dystrophin-alternative transcripts and what role they might play in normal development and the phenotype of Duchenne dystrophy. As more human disease genes are mapped to specific human chromosomes, we believe that our candidate gene approach, which builds upon knowledge of abnormal dystrophin, should help in the rapid characterization of these diseases.

Structural Studies on DNA-Replication Enzymes, src-related Oncogene Products, and Oxidoreductases

John Kuriyan, Ph.D.—Assistant Investigator

Dr. Kuriyan is also Associate Professor of Molecular Biophysics at the Rockefeller University. He received a B.S. degree in chemistry from Juniata College, Pennsylvania, and a Ph.D. degree from the Massachusetts Institute of Technology, where he worked with Gregory Petsko and Martin Karplus on the dynamics of proteins. He continued with Martin Karplus at Harvard University for a year and then moved to the Rockefeller University as a University Fellow. He is also a Pew Scholar in the Biomedical Sciences.

OUR interests are in understanding the structural determinants of protein function. To this end we carry out x-ray diffraction experiments and computer simulations aimed at characterizing the three-dimensional structures of proteins. We apply the knowledge gained from these studies toward the design of mutations and inhibitors to modify protein activity, and seek, from the particular cases in hand, to extrapolate to families of related proteins. Our current efforts are focused on three main areas: DNA replication, *src*-related oncogene products (tyrosine kinases), and redox enzymes and oxidative stress.

Structure of the β-Subunit of DNA Polymerase III

DNA polymerases are enzymes that duplicate the information content of DNA by catalyzing the template-directed polymerization of nucleic acids. Polymerases that are involved in chromosomal replication, such as DNA polymerase III (PolIII) of *Escherichia coli*, are distinguished by their high processivity—i.e., their rapid replication (1,000 bases/second) of very long stretches of DNA without dissociation. Processivity is conferred upon the enzyme by one of its subunits, β (processivity factor), which acts to clamp the polymerase onto DNA. The β-subunit binds very strongly—indeed, cannot be easily separated from circular DNA. It has been shown, however, to slide freely along duplex DNA, consistent with its role as a clamp that tethers the polymerase core to the template and advances with the polymerase during replication.

We have crystallized the β-subunit and determined its structure by x-ray diffraction to 2.5-Å resolution (Xiang-Peng Kong, Rockefeller University; Michael O'Donnell [HHMI] and Rene Onrust, Cornell University Medical College). Two molecules of the β-subunit are tightly associated to form a donut-shaped structure that forms a closed ring around duplex DNA. An unexpected feature of the structure is that it is highly symmetric. Each monomer consists of three domains of identical chain topology. Each domain is roughly twofold symmetric in its architecture, with an outer layer of two β-sheets providing a scaffold that supports two α-helices. Replication of this motif around a circle results in a rigid molecule with 12 α-helices lining the inner surface of the ring and with six β-sheets forming the outer surface.

The high symmetry of the structure is well suited to interact with the cylindrically symmetric DNA duplex, and the hole in the middle of the ring is large enough to accommodate either A or B forms of DNA with no steric repulsion. Although the overall charge on the protein is negative, a number of basic side chains on the inner surface generate a positive electrostatic field localized in the hole of the donut, precisely where the negatively charged phosphate backbone of DNA is expected to be. We are carrying out molecular dynamics simulations of a duplex-DNA and β-subunit complex, which is expected to provide information not readily accessible by x-ray crystallography. Other extensions of this project include crystallization attempts on the eukaryotic processivity factor PCNA (proliferating cell nuclear antigen), other subunits of PolIII, and other proteins acting at the replication fork.

Oncogene Products Related to *src*

Tyrosine kinases such as the product of the *src* gene play a central role in signal transduction pathways in the cell, and abnormalities in their function can result in dramatic changes in cellular differentiation and life cycles. Our approach is to work on these proteins in isolation, and we are collaborating with Hidesaburo Hanafusa (Rockefeller University) and Marilyn Resh (Sloan-Kettering Institute) to crystallize various functional domains of these oncogene products and their cellular equivalents. Initial success has been obtained for the *src* SH2 domain, which is responsible for binding to proteins containing phosphorylated tyrosine residues. Starting from *src*-containing plasmids (provided by Resh), Dorothea Kominos and Gabriel Waksman in our laboratory have cloned, overexpressed, purified, and crystallized SH2. The crystals diffract to 2.5-Å resolution, and the x-ray structure determination

of a complex of SH2 with a phosphotyrosyl peptide has been completed at 1.5-Å resolution.

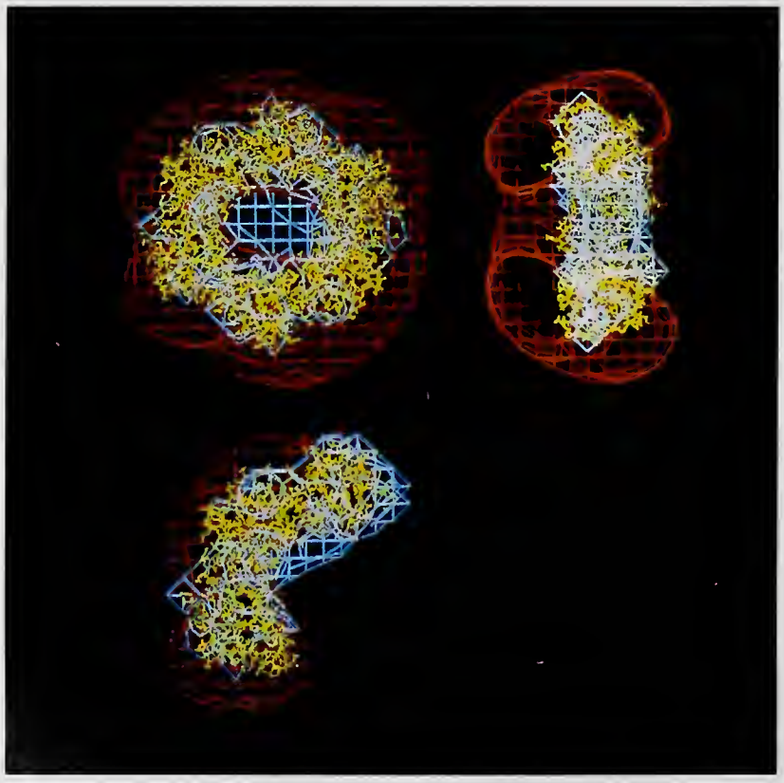
Redox Proteins and the Transcriptional Response to Oxidative Stress*

We have recently solved and refined the three-dimensional structures of two related redox enzymes, thioredoxin reductase and trypanothione reductase. Crystallographic investigations of enzyme-substrate complexes are now in progress (G. Waksman). Work has also begun on a new member of the disulfide reductase family. Proteins related to thioredoxin have been implicated in the process by which disulfide-containing proteins fold up rapidly without scrambling their disulfide pairings.

James Bardwell, Karen McGovern, and Jon Beckwith (Harvard Medical School) have recently identified an *E. coli* protein that is required for the correct folding of disulfide-containing proteins *in vivo*. This 21-kDa protein, the product of the *dsbA* gene, has no detectable sequence similarity to any known protein except for a short stretch of amino acids with similarity to the active site of thioredoxin (including the redox-active disulfide bond). The dsbA protein has been purified in our laboratory, from cells provided by Bardwell and Beckwith, and single crystals that diffract to 2-Å resolution have been obtained. The determination of the three-dimensional structure is under way, and we are also pursuing the construction of site-specific mutants at the active site (Jennifer Martin).

We are interested in determining the mechanism of the bacterial oxidative stress sensor OxyR. The OxyR protein responds to oxidizing agents by elevating the expression of such redox enzymes as catalase, superoxide dismutase, and glutathione reductase. The intact OxyR molecule proved to be unstable in the absence of DNA. In collaboration with Gisela Storz (National Institutes of Health), Scott Robertson in our laboratory has cloned, overexpressed, and purified the regulatory domain alone, and its crystallization is being pursued.

* This work is being supported by a grant from the National Institutes of Health.



Representations of the structure of β-subunit—the sliding clamp—of DNA polymerase III of Escherichia coli*, determined by x-ray crystallography. The atomic structure is shown in yellow and its electrostatic potentials in red (negative) and blue (positive). The top two images are different views of the dimer, and the bottom is a single monomer. The electrostatic potential generated by the protein helps stabilize the dimer and its interactions with the DNA molecule (not shown), which threads through the clamp formed by the protein.*

From Kong, X.-P., Onrust, R., O'Donnell, M., and Kuriyan, J. 1992. Cell *69:425–437. Copyright © 1992 by Cell Press.*

Molecular Analysis of Down Syndrome

David M. Kurnit, M.D., Ph.D.—Investigator

Dr. Kurnit is also Professor of Pediatrics and Human Genetics at the University of Michigan Medical School. He received his M.D. and Ph.D. degrees in cell biology from Albert Einstein College of Medicine. He did his internship and residency in pediatrics at the University of Pittsburgh and his fellowship in medical genetics at the University of Washington. After seven years as a Harvard Medical School faculty member at Children's Hospital, he joined the University of Michigan Medical Center.

WE are primarily interested in the etiology and pathogenesis of Down syndrome.

Why does the risk of having an offspring with Down syndrome increase dramatically as women age? There are two competing theories. The "older egg" model states that the older a woman becomes, the more abnormal (aneusomic) eggs she produces. The "relaxed selection" model states that as women age, they lose the ability to abort abnormal conceptuses. To determine the correct hypothesis, or whether a combination of the two applies, it is necessary to study a large number of Down syndrome cases.

Given our important (and unexpected) finding that 94 percent of nondisjunction errors are maternal in origin, resolution of the above controversy requires an ability to distinguish the stage of maternal cell division (meiosis I or II) in which the error occurs. To accomplish this, we will analyze four distinct probes (H6-5-6, 126-4-1, D21S120, and GT14) that detect polymorphic oligo d(A,C) sequences in the pericentromeric region of chromosome 21. Conditions and sizes have been achieved that allow each of these probes to be multiplexed in a single lane without co-interference. Using four different fluorescent markers, it is feasible to analyze electrophoretically in a single gel lane a molecular weight standard and a family consisting of father, mother, and their child with Down syndrome. Pouring the gels is the labor-intensive procedure in this protocol, and the ability to multiplex three family members and four polymorphisms in a single lane will yield a 12-fold savings in gel-pouring effort. This savings is made worthwhile by the fact that we must study at least 500 families to comprehend the biological basis for nondisjunction. This work was also supported by a grant from the National Institutes of Health.

To study pathogenesis, we first isolated genes encoded by chromosome 21, using a novel recombination-based methodology. Genomic fragments isolated from yeast artificial chromosomes (YACs) that map to chromosome 21 are cloned into a plasmid vector with the genetic marker *sup*F. RNA isolated from a variety of 20-week fetal human tissues was used to construct bacteriophage λ cDNA libraries, which are then infected into cells harboring a *sup*F plasmid carrying a nonrepeated genomic fragment on chromosome 21. (Using tissues recently obtained from earlier abortus specimens, we will now be able to make complex cDNA libraries from small amounts of these tissues by polymerase chain reaction techniques.) If any member of the genic cDNA library in λ shares homology with the genomic DNA fragment in the plasmid, then recombination mediated by that homology will ensue.

Following recombination between the bacteriophage and the plasmid, selection for bacteriophages carrying a given plasmid with *sup*F will result in selection for bacteriophages carrying a cDNA that is homologous to a genomic DNA cloned in the plasmid. In other words the system is designed to select for genomic sequences that are transcribed. The system is designed to stand alone or to interdigitate with the genomic initiative as it proceeds. In the latter case, as sequencing detects open reading frames that represent candidates for transcription, the recombination-based assay is designed to delineate analytically the tissue and timing of transcription and to result preparatively in isolation of the transcribed sequence.

The bulk of cDNA (gene) libraries are contaminated with small amounts of ubiquitous DNA sequences from plasmid pBR322. To circumvent this problem, we have cloned the *sup*F gene into an R6K-derived plasmid that lacks homology with pBR322 and have made other improvements that should increase the applicability of this procedure. We should now be able to screen a wide variety of extant cDNA libraries for transcription by chromosome-specific elements.

This systematology applied in model experiments has demonstrated that we can perform both selection and counterselection appropri-

ately. We are currently analyzing transcribed sequences on distal chromosome 21q (where the Down syndrome phenotype maps) and have isolated multiple candidates. The technique can also be applied to determine the timing and tissues of transcription of likely-to-be-transcribed sequences, as determined by other methodologies. We will collaborate with researchers performing these other techniques to make such determinations.

Replication and Pathogenesis of RNA Viruses

Michael M.-C. Lai, M.D., Ph.D.—Investigator

Dr. Lai is also Professor of Microbiology and Neurology at the University of Southern California School of Medicine, Los Angeles. He obtained his M.D. degree from National Taiwan University. He studied retroviruses with Peter Duesberg at the University of California, Berkeley, where he obtained his Ph.D. degree in molecular biology and continued for postdoctoral work.

MEDICAL history is marked by extraordinary successes against viral infections, but it is also punctuated by the continual emergence of new viruses. Since viruses, in general, contain very limited genetic information, they must rely on host cells for their own growth. How they cause diseases and how they continue to flourish in nature are not only interesting subjects in themselves but offer a lesson in the everyday workings of normal cells. Our laboratory is interested in RNA viruses, replicating entities in which RNA, in contrast to DNA, is the genetic material. We are exploring how these viruses replicate and cause diseases.

One of those we are studying is the coronavirus, named for its similar appearance to the corona of the sun. The virus causes the common cold in humans and a variety of gastrointestinal and respiratory problems in animals. It also causes symptoms very similar to those of multiple sclerosis, thus providing a model system for studying this disease. The virus has an RNA genome of 31,000 nucleotides, which is the longest known stable RNA. We are interested in learning how this unusually large RNA expresses its genes and maintains its genetic stability, despite an overwhelmingly high frequency of error in RNA synthesis. We have recently determined the complete sequence of the genome, giving us a glimpse of how the viral genes express themselves.

The virus utilizes a novel RNA synthesis mechanism, a discontinuous process that fuses a leader RNA to a gene located some distance from it. This unusual mechanism allows the leader RNA to control the expression of viral genes. We have recently found that the synthesis of one of the viral surface proteins indeed can be altered by minor changes in the leader RNA sequences, resulting in the variation of the viral pathogenicity. Furthermore, this variation can be observed during natural viral evolution. These observations suggest that the enzyme catalyzing coronaviral RNA synthesis is unusual, which is, indeed, suggested from the sequence of the gene encoding the enzyme. Our laboratory is investigating this novel RNA synthesis mechanism. A research grant from the National Institutes of Health provides support for this part of the research program.

Another unusual characteristic of coronavirus RNA has been revealed in our findings: it can undergo genetic exchange (RNA-RNA recombination) at an extraordinarily high rate. RNA-RNA recombination was previously thought to be a rare event in nature. We demonstrated, however, that it occurs readily between coronaviruses. This recombination can take place almost anywhere in the RNA genome, both in tissue culture cells and during animal infections. We recently succeeded in establishing that with the coronavirus, recombination can occur not only between two replicating viral RNAs but between viral RNA and a piece of an RNA fragment existing inside the cell, thus providing a model system for RNA recombination between viral and cellular RNAs. This type of recombination is one of the mechanisms by which some animal viruses become pathogenic.

We have demonstrated that a recombinant virus could become a predominant viral population under certain conditions, replacing the parental virus by a simple natural selection process. Thus recombination represents a powerful evolutionary tool for RNA viruses. From the standpoint of viral biology, RNA recombination may be the genetic mechanism by which coronaviruses weed out defective RNA sequences generated by errors in RNA synthesis. Coronavirus is thus able to maintain an RNA genome larger than was thought theoretically possible. RNA recombination has now been demonstrated in many different viruses, suggesting its important role in virus evolution in nature.

This genetic phenomenon also has an important implication in vaccine development for diseases such as AIDS (acquired immune deficiency syndrome), since genetic exchanges between viruses may lead to genetic instability of live, attenuated virus vaccines. We are continuing to study the RNA recombination mechanism and attempting to use it as a genetic tool in determining how viruses cause diseases.

Another virus we are studying in our laboratory

is the hepatitis delta virus (HDV), a human hepatitis virus commonly associated with a severe form of the disease. HDV, by itself, does not infect humans, because it is defective and requires another viral agent, hepatitis B virus (HBV), to supply an essential envelope protein to infect liver cells. HDV has been shown to cause epidemics of fulminant hepatitis in many parts of the world. In the United States, it is prevalent among intravenous drug abusers.

The virus contains a circular, single-stranded RNA genome of only 1,700 nucleotides. It is the only animal virus with a circular RNA. This genome structure is reminiscent of a group of plant pathogens, viroids or virusoids, that cause a variety of plant diseases. Indeed, the similarity between HDV RNA and plant viroid RNAs goes beyond their circular RNA structure. There are several structural and biochemical features that suggest a close evolutionary relationship between HDV and plant viroid RNAs. Both RNA groups contain a "ribozyme" activity, in which the RNA serves as an enzyme that cleaves and ligates the RNA itself. Thus HDV RNA stands at a peculiar place in the evolutionary ladder: it may have been derived from a plant pathogen by recombination with a gene that gave it the ability to infect human cells and cause diseases. Our laboratory is studying the properties of this ribozyme activity. We have determined its structural and sequence requirements, which are distinct from those of other known ribozymes. Furthermore, we have begun to examine whether ribozyme activity is the same in infected cells as in the test tube. The data obtained thus far suggest the interesting possibility that some cellular factors could participate in the ribozyme activity. We are pursuing these factors.

One important difference between HDV RNA and plant viroid RNAs is the ability of the former to synthesize a protein, hepatitis delta antigen (HDAg), the HDV signature protein. HDAg is required for RNA synthesis. We have been studying this protein's properties and functions and have shown that HDAg is a phosphoprotein that resides in the nuclei of the cells. It can interact with itself to form a protein complex, and it can also interact with HDV RNA in a specific way. We have shown that HDAg utilizes a set of novel sequence motifs to allow itself to bind to RNA.

What is the role of this protein, HDAg, in HDV RNA synthesis?—an unusual synthesis because HDV RNA is so small (1,700 nucleotides) that it lacks the capacity to provide its own synthesizing enzymes. Therefore, HDV most likely borrows cellular enzymes to do the job, which is uncharacteristic of RNA viruses. Most RNA viruses have to make their own enzymes, since normal cells do not appear to have this type. We are studying how HDV modifies the cellular enzymes to perform this rather atypical RNA synthesis and how HDV RNA synthesis initiates this. It appears that HDAg participates in these processes.

HDV thus provides a perspective on viral strategies from the small end of the RNA spectrum. Our laboratory is studying one of the largest RNA viruses (coronavirus) and the smallest (HDV), which utilize different principles for viral replication. Our studies not only offer insights into how these viruses cause diseases but also into fundamental mechanisms of RNA synthesis and RNA evolution.

Molecular Biology of Human Papillomaviruses

Laimonis A. Laimins, Ph.D.—Assistant Investigator

Dr. Laimins is also Associate Professor of Molecular Genetics and Cell Biology at the University of Chicago. He received his Ph.D. degree in biophysics and theoretical biology from the University of Chicago. His postdoctoral research was done with George Khoury at the National Cancer Institute.

MY laboratory is studying the molecular biology of human papillomaviruses types 16 and 18. HPV-16 and -18 are the etiological agents of the many malignancies of the urogenital region, and in particular those of the cervix. More than 70 different types of HPV are known to infect cutaneous and mucosal epithelia in various body locations. Some types induce warts on the hands or soles of the feet, but are never found in genital lesions. Similarly, the viruses that induce warts in the genital region are never found on the hands or feet.

The majority of HPVs cause benign warts, and about one-third of the types are specific for genital epithelium. A subset of these viruses (HPV-16, -18, -31, and -51) infect the cervix and are strongly associated with the development of cervical cancer. Although the number of HPV-positive individuals has increased 10-fold in the last 10 years, effective monitoring procedures such as the PAP smear have limited the increase in cervical cancers in the United States and Europe.

Up to 20 percent of the population in these countries is infected by some type of genital HPV, particularly the non-oncogenic types 6 and 11. However, many cases of infection by HPV-16 and -18 without visible lesions have been reported. It therefore appears that although infection by HPVs is necessary for the development of cervical cancer, infection alone is not sufficient to induce malignancy. In low-grade lesions, or condylomas, viral DNA is present in the cell as episomes, and progeny viruses are produced. In contrast, in cervical cancers the viral genome is found integrated into the host chromosome, and no viral particles are synthesized. One model for progression from a low-grade lesion to a cancer suggests that integration is important in the development of the cancer and that it may act by the removal of a dominant inhibitor of transformation. Viral integration into the chromosome may thus be the first step of a multistep carcinogenic process. My laboratory is currently studying the molecular mechanisms by which HPVs contribute to this process.

A major impediment to the study of HPVs has been an inability to propagate these viruses in culture. Papillomaviruses are unusual in that their life cycle is tightly coupled to the differentiation program of epithelial cells. Although most viruses infect one cell type and undergo a productive infection in the same cell, HPVs infect basal epithelial cells, establish their genomes as episomes, and only replicate upon cellular differentiation, with amplification of viral DNA and expression of late genes. Inability to duplicate this differentiation process *in vitro* long prevented the successful propagation of virus. We have recently duplicated several features of these productive infections of HPV in culture, and current studies are directed at identifying how these differentiation-specific properties are controlled. This work is supported by a grant from the American Cancer Society.

My laboratory has found that HPV can immortalize epithelial cells derived from either human foreskin or cervix in tissue culture. Normally primary cells have only a limited life span *in vitro,* but the presence of the HPV E7 protein is sufficient to allow for unlimited growth in culture.

The loss of differentiation is usually a characteristic of many epithelial cancers, including those of the cervix. We have been studying how HPV viral proteins alter epithelial cell differentiation. In low-grade lesions in which infectious viruses are produced, these proteins only slightly alter epithelial differentiation. In cervical cancers, on the other hand, infected cells have lost all ability to differentiate.

We have also used a system that accurately mimics the differentiation properties of epithelial cells *in vitro*, to show that HPV sequences alter the ability of epithelial cells to differentiate. The morphological changes that are seen in this tissue culture system are very similar to those seen in low-grade cervical neoplasias. In tissue culture, our HPV cell lines quickly lose the ability to differentiate and develop the appearance of high-grade neoplasias or cancers. Two HPV genes, E6 and E7, seem to be required for this transformation process. We believe this loss of differentiation may be an *in vitro* model for the development of cervical cancer. In the future, we hope to utilize this model to identify other important factors involved in controlling the development of

malignancy. These observations also strongly support the etiological role of HPV in the development of cervical cancer. This work was supported in part by a grant from the National Institutes of Health.

My laboratory has been able to identify factors involved in the cell type–specific regulation of viral gene expression and replication. We have identified viral enhancer sequences that are responsible for regulating viral expression in maturing keratinocytes. The function of one of these sequences is dependent on the action of a cellular protein that is only found in epithelial cells. We have designated this protein keratinocyte-stimulating factor, KRF-1, and are currently cloning the gene that encodes it.

Eventually we will examine how the action of cellular transcription factors is altered by the presence of viral transforming genes. We have shown that the products of both E6 and E7 can indirectly regulate the transcription of a variety of genes. In addition, we believe that a major determinant of the tissue spectrum of papillomaviruses is specified through the regulation of transcription. Understanding the mechanisms of tissue-specific expression will facilitate our analysis of why different HPV types only induce lesions in particular kinds of epithelial cells. Our studies on the biology of HPV should provide information on mechanisms of transformation and tumor progression, as well as tissue- and differentiation-specific expression.

Genetic Studies in Cardiovascular Disease

Jean-Marc Lalouel, M.D., D.Sc.—Investigator

Dr. Lalouel is also Professor of Human Genetics at the University of Utah School of Medicine. He obtained a medical doctorate, a master's degree in microbiology and genetics, and a doctorate of sciences in genetics at the University of Paris, France. He furthered his training as a postdoctoral fellow and a research associate with Newton Morton at the University of Hawaii and was Professor of Human Biology at the University of Paris before joining the faculty of the University of Utah.

COMMON cardiovascular disorders such as coronary artery disease and essential hypertension exhibit a familial tendency. Such broad clinical classes represent complex etiological entities, where many genes and environmental determinants are likely to be involved. Multiplicity and heterogeneity in causation stretch the ability of genetic methods to the limit of their investigative power. The consideration of biochemical parameters can reduce this degree of complexity. Before the global picture can be comprehended, however, details of the landscape will need to be scrutinized first. This perspective is highlighted in two examples from our laboratory. Work on lipoprotein lipase was led by Mitsuru Emi and Akira Hata, while work on hypertension was led by Richard Lifton.

Abnormal lipoprotein concentrations in plasma are commonly observed in the relatives of patients with early coronary disease, yielding various patterns of hypercholesterolemia and/or hypertriglyceridemia within families. Such complex phenotypes are thought to result either from the variable expression of a single-gene defect or from the independent contribution of two or more genes. The contribution of such genes to the clinical expression of hyperlipidemia is further blurred by the fact that hormonal influences, diet, and habitus exert major influences on the regulation of lipid metabolism. A severe form of familial hypercholesterolemia, which accounts for about 5 percent of myocardial infarction, has been linked to molecular defects of a lipoprotein cellular receptor. More than 90 percent of dietary fat is hydrolyzed by lipoprotein lipase (LPL) in an initial step controlling its delivery to peripheral tissues. The enzyme, secreted by adipocytes and muscle cells, acts at a distance from its site of synthesis, becoming anchored to the luminal surface of capillaries by an ionic interaction with heparan sulfate. In the presence of a specific cofactor, apolipoprotein C-II, the enzyme hydrolyzes triglycerides of intestinal or hepatic origin by binding to the surface of circulating lipoproteins, thereby releasing fatty acids for uptake in the tissues where they can be used as fuel or reesterified for storage. Defective functional enzyme is the diagnostic feature of a rare recessive chylomicronemia syndrome, familial LPL deficiency.

This condition is characterized by massive chylomicronemia in the fasting state, episodes of abdominal pain, life-threatening acute pancreatitis, and eruptive xanthomas. By investigating the relatives of a homozygous subject, we showed previously that heterozygotes for such molecular defects tend to express a common form of hypertriglyceridemia. However, when we screened unrelated subjects for molecular defects of the LPL gene, we found that mutations of this gene probably account for 3–5 percent of such hyperlipidemias. This situation, similar to receptor defects in familial hypercholesterolemia, again stress that most forms of hyperlipidemia are of heterogeneous origin.

We and others have now identified a host of mutations in the LPL gene of subjects with LPL deficiency, including a large number of simple amino acid substitutions. Such mutations may provide natural, specific probes of functional domains of the enzyme. When they are superimposed to the known three-dimensional structure of the homologous enzyme, pancreatic lipase, it becomes clear that they are spread over the protein's first folding domain, which includes the triad of residues directly involved in catalysis (see figure).

We reproduced four such mutations *in vitro*, expressed them in cultured cells, and analyzed the corresponding mutant proteins. A common feature of these mutations is that they affected the assembly and stability of the two identical subunits characterizing the active enzyme. Hence the majority of these mutations lead to loss of catalytic activity through rather nonspecific mechanisms.

To probe other domains of the enzyme, it became clear that we could not rely solely on naturally occurring mutations. LPL binds to heparan sulfate, a member of a family of ubiquitous and abundant complex polysaccharides. Interactions of this nature play a role in many biological processes, such as lipolysis, hemostasis, cell adhesion, cell proliferation, or angiogenesis. The pro-

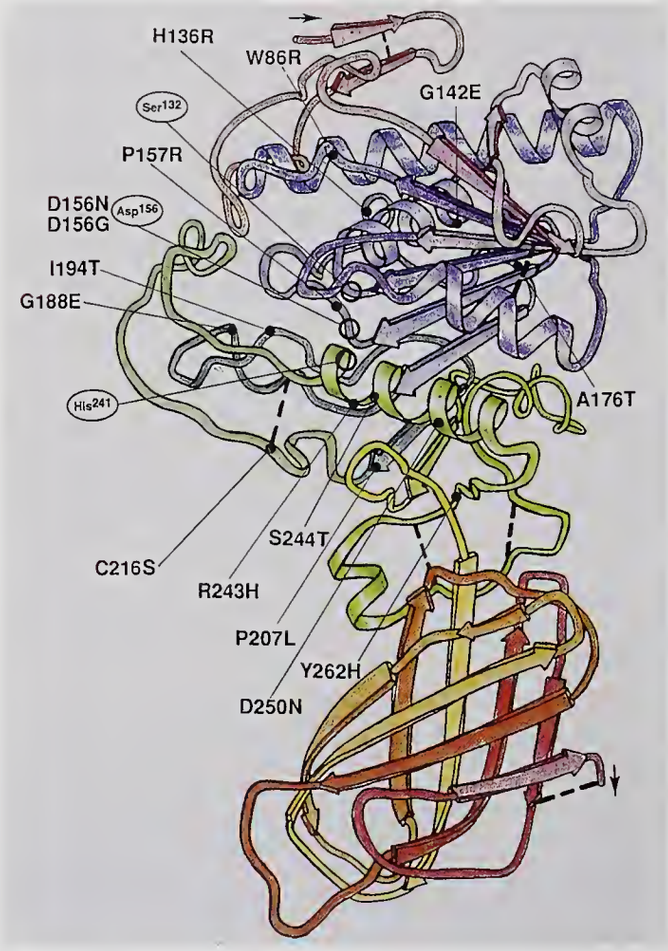
tein determinants involved in these interactions are poorly characterized, but the high density of negative charges on heparan sulfate indicates that the protein domains should include positively charged residues.

After generating many artificial mutations at such positions, we have produced the corresponding proteins by expression in cultured cells and analyzed their affinity for heparin. We have identified six basic amino acids that mediate the high affinity of the active, dimeric enzyme with heparin. They form a domain of unique structure, which we are attempting to characterize further.

Essential hypertension presents a yet greater challenge, for little biochemical insight is usually available. Glucocorticoid-remediable aldosteronism (GRA) is a rare autosomal dominant disorder marked by severe hypertension and hyperaldosteronism with high levels of abnormal adrenal steroids. All these manifestations can be corrected by the administration of glucocorticoids.

Aldosterone is a steroid involved in the regulation of the balance of sodium and potassium ions and produced in the zona glomerulosa of the adrenal gland under the primary control of the renin-angiotensin system. By contrast, glucocorticoids exert their effects on carbohydrate metabolism, are produced by the zona fasciculata of the adrenals, and are regulated by the adenohypophyseal hormone adrenocorticotropin (ACTH). Their synthetic pathways share several enzymes, including 11β-hydroxylase. Aldosterone synthesis, however, requires a unique enzymatic step catalyzed by aldosterone synthase, which is normally expressed only in the zona glomerulosa.

In a subject with GRA, we found that unequal crossing over between 11β-hydroxylase and aldosterone synthase, in close proximity on chromosome 8, had created a new chimeric gene comprising regulatory sequences of 11β-hydroxylase and sequences responsible for the catalytic specificity of aldosterone synthase. This observation explains the ectopic production of aldosterone in the adrenal tissue responsible for synthesis of glucocorticoids, and the corresponding hormonal control observed in GRA. Although this finding provides a clear interpretation of the complex physiology of a rare form of hypertension, its significance for our understanding of the more common forms of essential hypertension remains to be determined.



Mutations of human lipoprotein lipase that impair the enzyme's ability to regulate lipid metabolism are shown superimposed upon a schematic rendering of the backbone of a homologous enzyme, human pancreatic lipase. The three circled residues define the triad directly involved in catalysis.

From Lalouel, J.-M., Wilson, D.E., and Iverius, P.-H. 1992. Curr Opin Lipidol *3:86–95.*

Structure and Replication of Influenza Virus and Paramyxoviruses

Robert A. Lamb, Ph.D., Sc.D.—Investigator

Dr. Lamb is also John Evans Professor of Molecular and Cellular Biology and Professor of Microbiology-Immunology at Northwestern University. He received his undergraduate degree reading biochemistry at the University of Birmingham, England, and his Ph.D. and Sc.D. degrees from the University of Cambridge. He came to the United States to do postdoctoral work with Purnell Choppin at the Rockefeller University, where he later became a faculty member. Ten years ago he joined the faculty of Northwestern University.

ANIMAL viruses provide a unique tool with which to study the complex biochemical processes involved in the biosynthesis and maintenance of eukaryotic cells. Our laboratory is investigating the molecular structure and the mechanism of replication of two enveloped RNA viruses, influenza virus and the paramyxovirus SV5.

Influenza virus causes important diseases in humans and animals. It has tremendous socioeconomic consequences, for influenza continues to occur in regular epidemics and occasional pandemics and is a leading cause of morbidity and mortality. Paramyxoviruses cause many biologically and economically important diseases of humans and lower animals. Besides SV5, these viruses include parainfluenza virus types 1–5, mumps virus, measles virus, canine distemper virus, Newcastle disease virus of chickens, and rinderpest of cattle.

We have been elucidating the wide range of mechanisms used by these RNA viruses to maximize the amount of encoded protein in their compact genomes. We have identified overlapping reading frames, splicing of mRNAs, the use of bicistronic mRNAs, transcriptional stuttering to add nontemplated nucleotides to an RNA transcript (and hence yield a separate mRNA), and a coupled stop-start translation of tandem cistrons.

Influenza virus and SV5 were selected for study not only because of their importance as the causative agents of major diseases but also because they provide excellent models for examining a variety of properties of integral membrane proteins. Since these proteins are the major antigenic determinants of the viruses, knowledge about their structure should enhance our understanding of how they act as immunological targets, thus aiding in developing new vaccines. In addition, some of the biochemical activities of the influenza virus are specialized to the virus, making them attractive as points of intervention in the virus life cycle to which rationally designed therapeutic agents can be developed. We are analyzing biochemical properties of the viral glycoproteins.

We are also investigating the mechanism by which integral membrane proteins are transported to the cell surface in the exocytotic pathways and are internalized from the surface by the endocytotic pathways. We are studying the seven integral membrane proteins encoded by influenza virus and SV5—three of which were discovered in our laboratory—because they provide a diverse group of membrane proteins that span the cell membrane once.

Virus Cation Channels

Influenza virus protein M_2 is a small (97-residue) type-III integral membrane protein that forms a disulfide-linked tetramer. The sensitivity of influenza virus to the drug amantadine hydrochloride, the coupling of antiviral action to the M_2 transmembrane domain, and the premature acid-induced conformational change in the viral hemagglutinin in the presence of the drug suggest that M_2 is an ion channel, that it is essential for virus uncoating in secondary endosomes, and that it can alter the intracellular pH of the trans Golgi network. In collaboration with Lawrence Pinto, Northwestern University, we have tested the M_2 protein for ion channel activity by injecting M_2 mRNA into *Xenopus* oocytes and measuring surface currents with a two-electrode patch-clamp apparatus. We have shown that expression of the protein is associated with an ion channel activity selective for monovalent ions.

Amantadine hydrochloride significantly attenuated the inward current induced by hyperpolarization of oocyte membranes, and mutations in the M_2 membrane-spanning domain that confer viral resistance to amantadine produced currents that were resistant to the drug. Thus we have provided direct data on the antiviral drug's mechanism of action.

The M_2 protein does not have the molecular structure of most ion channels cloned to date. We had to perform a large number of experiments to eliminate the possibility that the M_2 protein was not a regulator that activates a normally silent channel endogenous to oocytes. Our analysis of distinguishing characteristics of the currents pro-

duced by various M_2 proteins with changes in the transmembrane domain suggests that the domain forms the channel pore, and that the M_2 protein is therefore a channel per se.

Our experiments have also indicated that the M_2 protein channel activity is activated by low pH, suggesting that the channel is only switched on in endosomes and the trans Golgi network—intracellular compartments with lowered pH. We are currently beginning a detailed structure-function analysis of this channel to characterize further the residues involved in drug sensitivity and ion specificity.

We have also shown that the influenza B virus NB glycoprotein, which has a similar overall structure but no obvious amino acid homology to the influenza A virus M_2 protein, has ion channel activity when expressed in oocytes. However, the conductance of the NB ion channel activity is specific for chloride ions and is regulated by calcium ion concentration.

The proposed pivotal role of ion channels in the replicative cycle of the influenza viruses suggests that the proteins present an important target for a point of intervention by drugs (in addition to amantadine) in the prophylaxis and therapy of virus infections. We are currently testing whether other viral proteins with similar structures to M_2 and NB have ion channel activities—e.g., the paramyxovirus SH protein, the *vpu* integral membrane protein of the human immunodeficiency virus I, and the C3 protein of the coronavirus avian infectious bronchitis.

Intracellular Transport of Glycoproteins

To elucidate the rules that govern protein orientation in the lipid bilayer, we are examining how polypeptides are initially inserted into the endoplasmic reticulum (ER) and are determining the signals necessary for the protein-bilayer interaction. One of the major factors is the presence of positively charged residues flanking the hydrophobic membrane-spanning domain to retain a region of the protein in the cytoplasm. We have also been focusing on the factors and signals needed to fold the primary polypeptide chain once it has been translocated across the membrane of the ER.

The cellular glucose-regulated protein GRP78-BiP is a member of the HSP70 stress family of gene products and is a resident component of the ER, where it is thought to play a role in the folding and oligomerization of secretory and membrane-bound proteins. GRP78-BiP also binds to malfolded proteins and this may be one mechanism for preventing their intracellular transport. During folding, the SV5 hemagglutinin-neuraminidase (HN) glycoprotein specifically and transiently associates with GRP78-BiP. This complex formation can only be detected prior to oligomerization of the immature HN molecules to form the native tetramer, suggesting that GRP78-BiP acts as a chaperone to promote correct folding of the molecule.

Internalization and Degradation of Glycoproteins

The SV5 HN glycoprotein is extensively internalized from the virus-infected cell surface and degraded in lysosomes. We are making an extensive study of the mechanism of HN internalization. This is of considerable interest because HN lacks an aromatic amino acid in its cytoplasmic tail that has been found necessary for the internalization of several well-characterized receptor molecules internalized by the clathrin-coated vesicle pathway. Examination of chimeric molecules constructed between HN and another type-II integral membrane protein that is not internalized, influenza virus neuraminidase, suggests that the HN transmembrane domain signals internalization from the cell surface and specifies targeting to lysosomes.

Some aspects of research in our laboratory on influenza virus and paramyxoviruses are supported by grants from the National Institutes of Health.

Cancer and Genetic Modification

Philip Leder, M.D.—Senior Investigator

Dr. Leder is also John Emory Andrus Professor in the Department of Genetics at Harvard Medical School. He received his M.D. degree from Harvard Medical School. He has also received three honorary D.Sc. degrees. Dr. Leder held several positions at NIH before returning to Harvard. His many honors include the Albert Lasker Medical Research Award, the National Medal of Science, and the Heinekin Prize awarded by the Royal Netherlands Academy of Arts and Sciences. He is a member of the National Academy of Sciences.

THE growth of cells in an organism is far too delicate a process to be left to chance. Rather, as with all biologic processes, it is subject to a very stringent set of rules that are programmed into the makeup of the organism. The basis for the control of growth is genetic. The genes set the parameters that allow, say, the liver to take the shape it does or the kidney to assume its particular size and function. Genes establish the rules whereby an organ grows in an orderly fashion and reaches a prescribed and limited size. Thus growth can proceed so far but no farther, attaining a programmed equilibrium compatible with life.

Cancer as a Disease of Genes

Cancer is a profound disorder of cell growth. The delicate balance established by a genetically encoded program is overturned. Instead of reaching an equilibrium, the cancer cell no longer responds to signals that would limit its ability to divide. It is out of control, and its unlimited growth has profoundly dangerous consequences for the organism.

Over the past decade or so, it has become increasingly clear that many cancers can be accounted for, at least in part, by damage occurring to genes that encode the rules for control of cell growth. Genetic damage, or mutation, can inactivate a gene or cause it to function at the wrong time or in the wrong place or, indeed, even cause it to make the wrong product. The set of genes whose mutation can give rise to cancer is often just those that normally regulate cell growth. Geneticists refer to the damaged genes that contribute to the development of malignancy as oncogenes (from the Greek *ονκοσ*, or tumor).

Transgenic Mice and the Genetic Basis of Cancer

For some time my colleagues and I have been interested in genes that control cell growth. Our work has been considerably advanced by the technique of introducing active oncogenes into the hereditary makeup of special strains of laboratory mice. Called "transgenic," such mice carry oncogenes created in the laboratory, pass on these cancer-causing genes to offspring, and therewith transmit a strong tendency to develop cancer. Thus, in many ways, transgenic mice become useful models of human malignancy.

For example, we have designed specific mice that develop cancer of the breast and others that develop cancer of the blood cells—specific leukemias and lymphomas. Some of these mice even develop benign prostatic hypertrophy, a condition that affects up to 85 percent of men by the eighth decade of life. These experiments have taught us that certain cancers can be caused by specific oncogenes and that many, but not necessarily all, cancers are the result of a collaboration between two or more oncogenes. This suggests that cancer is often a "multihit" process, one that requires several activating events.

A Binary System for Activating and Silencing Transgenes

During the past year we have extended the power of transgenic technology by creating a system that gives us much better control over the transgene we have introduced. For example, we often introduce genes that dramatically increase the incidence of certain cancers in our mice. Cancers obviously influence the ability of our animals to pass their genes on to succeeding generations, as such genes often preclude survival. To overcome this problem and to assure that no more cancer-prone mice are produced than we need for our experiments, we have designed a binary system in which "target" genes can be held in an inactive state in one line of mice and become active only when the mice are bred to a second line that carries an "activator" gene. The system is suggestive of epoxy cement that is held in two tubes, the contents of which must be mixed to become functional.

The binary system has the further advantage that it can be "multiplexed," or used in a variety of combinations, such that a target can be one of many different transgenes that is in turn combined with one or more activator genes. For example, the activators could specify expression of a

target in many different organs at different times. A further interesting aspect of this system is that the unity of biology permits elements of the system that have been derived from the regulatory machinery of a simple unicellular organism, yeast, to be used quite readily in mammals.

Host Defenses Against Cancer

While transgenic mice are very useful in analyzing the action of oncogenes, they are also useful in exploring the host defense mechanism mobilized to prevent the development and spread of cancer. The immune system is one of the organism's chief instruments against the spread of infectious disease and for the rejection of foreign tissues. For instance, the body's immune system must be neutralized to accomplish effective heart or kidney transplants. The role that the immune system plays in tumor rejection is poorly understood, but important discoveries in immunology indicate that the immune response is regulated by an array of hormone-like agents called lymphokines. These are released by cells of the immune system to influence the growth and development of other cells.

We are particularly interested in how lymphokines influence the host's response to cancer. In the course of this work, we have focused on the action of two particularly interesting lymphokines, IL-4 and IL-7. (These biologic response-modifying agents are frequently called IL, derived from "interleukin," an agent that mediates signals between white blood cells, or leukocytes.) Several cell-signaling functions have been recognized in IL-4. Although its precise role in the body has not been proved, IL-4 is suspected of playing a role in modulating the immune response. IL-7 is thought to be important for the orderly development of antibody-producing cells, the so-called B cells.

We have been able to show that both IL-4 and IL-7 are potent antitumor agents, acting to induce host defense mechanisms. Further studies of the action of IL-4 have allowed us to identify two particular cell types that may mediate this antitumor effect. One of these is the eosinophil, a cell involved in many allergic responses. Another is the macrophage, a scavenger cell concerned with cell-killing functions. Our most recent work has focused on identifying the active regions of the IL-4 molecule, with a view to distinguishing those structures required for antitumor activity from those that give rise to unwanted side effects.

New Directions: Distinct Genetic Contributions from Mothers and Fathers

Mendel's vision of genetics held that a particular gene carried by an organism behaves in exactly the same way whether it is inherited from the mother or the father. For the most part, this is true. Nevertheless, in mammals, the highest organisms, it has proved impossible to induce development artificially from the egg alone (a process known as parthenogenesis), although this is possible in many lower forms. Some experiments using transgenic mice have helped us to understand this phenomenon and why it is that, at least for mammals, mother and father (that is, both egg *and* sperm) are necessary for development of the offspring.

It turns out that a small number of genes are expressed differently if they are inherited from one or the other parent. In a particular example that we created, a transgene is only expressed if it is inherited from the father. The very same gene inherited from the mother is silent. We have now correlated this so-called "parentally imprinted" expression with a chemical modification of the gene whereby it is heavily altered (but not expressed) if inherited from the mother or not altered (but expressed) if inherited from the father. During the past year we have worked out the rules that govern the modification of this gene during embryogenesis. Our most recent work is directed toward identifying the encoded signal that evokes this parental effect.

From Molecular Biology to Therapy of Human Disease

Fred D. Ledley, M.D.—Assistant Investigator

Dr. Ledley is also Associate Professor of Cell Biology and Pediatrics at Baylor College of Medicine. He received his B.S. degree in physical sciences from the University of Maryland, College Park, and his M.D. degree from Georgetown University. He trained in pediatrics and medical genetics at the Children's Hospital, Boston, and Harvard Medical School. His postdoctoral research was conducted with David Baltimore at the Massachusetts Institute of Technology and with Savio Woo at Baylor College of Medicine.

SINCE its inception this laboratory has focused on genetic deficiency of the enzyme methylmalonyl CoA mutase (MCM) as a model for molecular genetic investigations of human disease. We have cloned and sequenced human and murine cDNAs for MCM, have mapped and cloned the genomic locus in mice and humans, and have described a series of mutations causing interesting and informative phenotypes of this disorder.

While continuing these studies, we have begun to address the technologies that will be essential in moving toward somatic gene therapy for MCM deficiency. We have focused our efforts on understanding the consequences of metabolic engineering of MCM activity in human cells, on developing methods for efficient manipulation and transduction of human hepatocytes, and on assessing methods for hepatocellular transplantation in large animals and humans.

Retroviral-mediated Correction of MCM *in Vitro*

We constructed a high-titer, amphotropic retroviral vector containing the full-length normal human MCM gene. MCM-deficient fibroblasts were transduced with this vector, and the efficiency of transduction was assessed by semiquantitative identification of the recombinant provirus. The restoration of MCM activity was measured by the ability of cells to metabolize [^{14}C]-propionate. The metabolic capacity of fibroblasts was restored to normal levels by transduction, even though only 10–30 percent of the cells were effectively transduced. No further increase in metabolic capacity was evident when MCM apoenzyme activity was increased above normal levels by varying transduction or transfection conditions. Transduction of normal fibroblasts or hepatoma cells increased MCM apoenzyme activity, but not the capacity for propionate metabolism, suggesting that other steps on these pathways are normally limiting.

Metabolic cooperation between cells was shown to increase the flux of propionate through subpopulations of metabolically competent cells. The capacity for propionate metabolism of hepatic cells was also found to be more than 10-fold greater than the maximal capacity of fibroblasts, suggesting that the phenotypic impact of metabolic engineering in hepatic cells would be greater than in other potential targets for gene therapy.

Transduction of Primary Human Hepatocytes

One strategy for targeting gene therapy to the liver involves harvesting and culturing hepatocytes from patients, transducing these cells *ex vivo* with retroviral vectors, and returning them to patients by hepatocellular transplantation. We demonstrated the feasibility of hepatocellular harvest from human liver segments preserved in UW (University of Wisconsin) solution, demonstrated proliferation of hepatocytes and preservation of differentiated hepatocellular functions in cells cultured in hormonally defined media, and established methods for transducing human cells with recombinant retroviral vectors.

The efficiency of transduction of human cells with amphotropic vectors (1–10 percent) was significantly less than that observed in various animal models. Vectors containing MCM were shown to constitute expression of recombinant mRNA from the 5′ long terminal repeat (5′LTR) promoter. Higher-efficiency transduction was obtained using retroviral vectors with xenotropic or gibbon ape *env* determinants. Ongoing studies are aimed at optimizing conditions for hepatocyte cultivation and transduction.

Hepatocellular Transplantation

One of the major factors limiting the application of *ex vivo* strategies for hepatic gene therapy is that hepatocellular transplantation has never been attempted in clinical practice. We have assessed the feasibility in large animal models, using a novel method for tracking transplanted cells. The cells are stained with a fluorescent dye, DiI. This dye is not metabolized or exchanged between cells, and engraftment can be identified in recipient tissues by fluorescent microscopy or flow cytometry.

Studies in mice and rats confirm previous reports that the efficiency of engraftment is relatively poor (less than 1 percent). We have studied autologous hepatocellular transplantation in baboons as a model for human anatomy and surgical methods. Transplantation resulted in constitution of approximately 5 percent of the host liver from the hepatocellular graft without complications.

We have explored the possibility of hepatocellular transplantation and gene therapy *in utero*. Heterologous hepatocellular transplantation was performed in fetal lambs at 80–85 percent gestation by infusion of cells into the umbilical vein. Flow cytometry of hepatocytes recovered from transplanted animals demonstrated that 1.5–4.5 percent of hepatocytes originated from the graft.

Transient Gene Therapy by *in Vivo* Gene Delivery

Other laboratories have reported that DNA coupled to asialoglycoproteins can be targeted to the liver *in vivo*, but that expression of these genes is transient (days). We are interested in the possibility that transient expression might be the most efficacious approach to treating MCM deficiency, which is characterized by periods of relative stability punctuated by life-threatening episodes of acidosis. We have demonstrated *in vivo* delivery of MCM to the livers of experimental mice using asialoorosomucoid-polylysine-DNA conjugates and are currently studying the consequences of this gene delivery. In particular, we are concerned with demonstrating that DNA is completely eliminated after the period of transient expression, since damage to chromosomal DNA from inadvertent integrations is one of the major potential risks. Vectors will be constructed with suicide sequences to eliminate integrated DNA.

Future Directions

Successful gene therapy requires attention not only to methods for gene delivery and gene expression, but consideration of the metabolic, cellular, clinical, and social consequences of genetic manipulation. We are attempting to establish a broad base of expertise and experience, using MCM deficiency as a model. This should enable rational development of clinical trials involving somatic gene therapy in the future.

Molecular Biology of Hormone and Drug Receptors in Health and Disease

Robert J. Lefkowitz, M.D.—Investigator

Dr. Lefkowitz is also James B. Duke Professor of Medicine and of Biochemistry at the Duke University Medical Center. He received his B.A. (chemistry) and M.D. degrees from Columbia University and clinical and research training at Columbia-Presbyterian Medical Center in New York, NIH, and Massachusetts General Hospital. As a molecular pharmacologist he has focused on the molecular structure and regulatory mechanisms controlling the function of the adrenergic receptors that mediate the actions of catecholamines. Dr. Lefkowitz has received numerous awards, including the Gairdner Foundation International Award. He is a member of the National Academy of Sciences.

OUR research program is concerned with the molecular properties and regulatory mechanisms that control the function of plasma membrane receptors for hormones and drugs under normal and pathological circumstances. Receptors are the cellular macromolecules with which biologically active substances (e.g., hormones, drugs, neurotransmitters, growth factors, viruses, lipoproteins) initially interact. Such receptors perform two essential functions: 1) They receive or bind these biologically active substances at the surface of the cell. 2) They transmit the substance's message into the cell, thus influencing its metabolic activity and function.

We have utilized the receptors for epinephrine (adrenaline) and related compounds, which are generally termed adrenergic receptors, as models for the study of receptors. Such receptors are found throughout the brain, heart, smooth muscle cells, and most other cells of the human organism. There are at least nine distinct subtypes of adrenergic receptors (α_1, α_2, β_1, β_2, etc.). These receptors interact not only with endogenous epinephrine and norepinephrine but with a variety of clinically important drugs.

We have studied these particular receptors for several reasons: they are more widely distributed than perhaps any other type of receptor, they are exemplary of each of the major biochemical pathways by which receptors are known to signal to the interior of the cell, and they are clinically and therapeutically significant. These receptors mediate physiological responses as diverse as changes in blood pressure, changes in heart rate and contractility, and alterations in the metabolism of glucose. Drugs that interact with these receptors are among the most effective agents used to treat various forms of heart disease, hypertension, asthma, shock, and depression. Research on these exemplary receptors has important implications for understanding hormone and drug receptor interactions generally and the mechanisms by which they are regulated.

Our research is focused on several intimately related goals. First, we wish to understand, in detailed molecular terms, the biochemical nature of the receptors. This is being accomplished by the application of recombinant DNA or gene-cloning techniques. These permit isolation of the genes for the receptors, which in turn permits their complete primary amino acid sequences to be deduced.

We have isolated the genes for all of the known adrenergic receptors, as well as a number of closely related receptors, and determined their complete sequences. Remarkably, the structures of these receptors are similar to each other and to that of the visual light receptor rhodopsin. This insight is helping to clarify the general mechanisms by which signals as divergent as a photon of light and a drug molecule like epinephrine elicit their characteristic biochemical and physiological responses. By varying the structure of the gene and hence the receptor protein that it produces, we can define which structural features of the receptor molecule determine its characteristic functions.

An unexpected result of the cloning of the known types of adrenergic receptor genes was the discovery of the genes for several novel subtypes of adrenergic receptors not previously known to exist. Knowledge of these new receptors opens the way for the development of new drugs with greater selectivity and fewer side effects. Such drugs might conceivably have applications in the treatment of such disparate diseases as hypertension, asthma, diabetes, and prostatism, or even as novel anesthetic agents.

A second current research goal is to elucidate the patterns of receptor regulation. One of the most important insights to come from our studies of receptors is that their properties are not fixed. Rather, the properties of the receptors are influenced by the hormones and drugs with which they interact, as well as by a variety of disease states.

There are important clinical implications of the ever-changing nature of the receptors. For example, this provides a basis for beginning to understand the phenomenon of drug tolerance or desensitization, the diminishing effect of drugs

over time. This phenomenon markedly compromises the therapeutic efficacy of epinephrine and many other drugs. When drugs like epinephrine combine with their receptors, they not only stimulate them but also produce changes that impair their function, thus leading to desensitization. These changes involve an actual loss of receptors from the cell surface (they move inside the cell where they cannot function) and a chemical change of those receptors remaining at the cell surface so that they function less effectively. With fewer functioning receptors present at their surface, cells are less able to respond to drugs or hormones.

Our recent research is increasing our understanding, in molecular terms, of how the receptors become functionally desensitized. We have discovered two new proteins that function to desensitize the receptors. The first is an enzyme, the β-adrenergic receptor kinase (βARK), which modifies the structure of the receptors by introducing a phosphate group when the receptors are stimulated. The second is the protein β-arrestin, which binds to the phosphorylated receptors and prevents them from acting. Both proteins are widely distributed at synapses throughout the central nervous system, suggesting their actions are not limited to the β-adrenergic receptors.

The implications of such fundamental research on receptors for clinical medicine are profound. Elucidation of the detailed structure of the receptors will allow the precise design of drugs that are more potent and specific. Unraveling of the molecular basis of desensitization will allow the development of strategies for interdicting the basic reactions that lead to loss of hormone and drug effect. An example is the design of specific enzyme inhibitors for βARK that could block the reactions leading to desensitization. Successful conclusion of such research may lead to methods for greatly prolonging and augmenting the therapeutic actions of diverse types of drugs.

Axis Formation and Germline Determination in Drosophila

Ruth Lehmann, Ph.D.—Assistant Investigator

Dr. Lehmann is also Associate Member of the Whitehead Institute of Biomedical Research, Associate Professor of Biology at the Massachusetts Institute of Technology, and Assistant Molecular Biologist at Massachusetts General Hospital, Boston. She received her M.Sc. degree from the University of Freiburg, where she worked with José Campos-Ortega on early neurogenesis in Drosophila. *She received her Ph.D. degree from the University of Tübingen, where she worked with Christiane Nüsslein-Volhard at the Max Planck Institute for Developmental Biology on the genetics of pattern formation in the* Drosophila *embryo. After postdoctoral training in Tübingen and at the MRC in Cambridge with Michael Wilcox, she joined the Whitehead Institute.*

HOW does a developing embryo know where to form a head and where to put a tail? In *Drosophila*, basic information about the "coordinates" of the embryo is supplied to the egg cell by the mother. Mutations in maternal genes have led to the identification of a small number of genes that are required for the establishment of anterior-posterior (head-to-tail) polarity. In these mutants the lack of a particular gene product in the mother is lethal to its progeny.

Three signals are required for the establishment of pattern along the anterior-posterior axis. The anterior signal controls the development of head and thorax, the posterior signal determines the abdominal region, and the terminal signal is required at both ends of the embryo. The anterior and posterior signals are localized to the respective poles (see figure).

We have concentrated on the dissection of the pathway leading to normal development of the posterior region. Nine maternal genes, called the posterior group, set the basic posterior pattern. These genes share the abdominal phenotype: homozygous mutant females produce offspring that lack abdominal segmentation. One gene, *nanos*, encodes the localized signal required for the development of the abdomen. Another gene, *pumilio*, regulates the activity of NANOS protein, while the remaining seven genes—*oskar*, *vasa*, *tudor*, *valois*, *staufen*, *cappuccino*, and *spire*—are required for the localization of *nanos* RNA to the posterior pole.

These seven genes are also required for germ cell formation. Embryos from mutant females lack the specialized posterior pole plasm that normally contains the polar granules. These embryos therefore fail to produce pole cells, the *Drosophila* germ cell precursors. Our molecular analysis of the posterior group is aimed at understanding how *nanos* and other RNA species become localized to the posterior pole and how the pole plasm, composed of RNA and protein, is assembled.

Synthesis and Function of the Abdominal Signal

Through genetic experiments as well as cytoplasmic transfers between wild-type and mutant embryos, we concluded that the specialized cytoplasm at the posterior pole is the source of an abdominal signal, and that embryos derived from females with mutant posterior group genes lack this signal in the abdominal region. The signal is encoded by *nanos*. We have isolated and cloned this gene and have shown that *nanos* mRNA is transcribed during oogenesis and becomes localized to the posterior pole plasm of the mature oocyte. The role of *nanos* as a signal for posterior pattern formation is demonstrated by the finding that injection of synthetic *nanos* mRNA into the anterior pole of early embryos leads to suppression of head formation and induction of a second abdomen in mirror image to the normal "posterior" abdomen.

Seven of the posterior group genes affect the localization of *nanos* mRNA to the posterior pole. Since these seven mutants also lack pole plasm, we can conclude that their abdominal segmentation defect is a consequence of a lack of localized *nanos* mRNA. We have identified sequences within the *nanos* mRNA required for its localization. In future experiments we will determine which of the posterior group genes are directly involved in tethering *nanos* mRNA to the posterior pole plasm.

Regulation of Posterior Activity

Several lines of evidence indicate that the *pumilio* gene is required to regulate the activity of NANOS protein. After fertilization of the egg, NANOS is translated from the localized mRNA and emanates anteriorly toward the abdominal region to form a gradient of protein concentration. Cytoplasmic transplantation experiments and analysis of the distribution of *nanos* RNA and NANOS protein in *pumilio* mutant embryos show that *pumi-*

lio does not interfere with *nanos* RNA localization or NANOS protein distribution, synthesis, and stability. We conclude that *pumilio* is required for optimal NANOS activity in regions of the embryo where NANOS protein concentrations are low.

We have cloned the *pumilio* gene and have shown that *pumilio* mRNA is localized to the posterior pole in early embryos. This observation is consistent with a possible interaction between the *pumilio* and *nanos* gene products. Localization of *pumilio* mRNA, like that of *nanos* mRNA, is disrupted in mutants that lack the specialized posterior pole plasm (see below). It is thus possible that the same mechanism of RNA localization acts on both mRNAs.

Assembly of the Pole Plasm

The remaining seven genes in the posterior group are associated with pole plasm defects. We are studying these genes from two perspectives. One is their role in pattern formation and body segmentation (see above); the other, their role in pole plasm assembly and germ cell formation (see below).

During the first hour of embryogenesis, the pole plasm appears as a distinct clear zone at the posterior end of the newly fertilized egg. Closer examination reveals that this zone contains specialized cytoplasm packed with mitochondria and numerous donut-shaped organelles, called polar granules, that do not occur anywhere else in the embryo. The zone of specialized cytoplasm coincides with the site of germ cell formation, and polar granules may therefore contain factors that control germ cell fate. Polar granules were first recognized in *Drosophila* embryos 30 years ago, but until recently little was known about their biochemical structure or function. Now the link between defects in abdomen formation and the absence of germ cells has provided a new avenue for research.

We have begun extensive studies of the pole plasm in mutant and normal embryos, using probes designed to recognize mRNA and proteins derived from the posterior group genes. We suspect that *oskar*, one of the genes required for abdominal and germline development in the *Drosophila* embryo, is an essential component of the pole plasm. We have isolated this gene. *In situ* hybridization to whole-mount embryos reveals that *oskar* mRNA distribution is restricted to the posterior pole of early embryos. Genetic studies suggest that pole plasm formation occurs in a series of steps, with each step dependent on the previous one, and that *oskar* is required early in the assembly pathway. Future experiments will address questions about the role of *oskar* mRNA and protein for pole plasm formation and for localization of the abdominal signal.

We are optimistic that new information about the assembly and composition of the pole plasm and polar granules will lead to better understanding of their functions in the early embryo. Their primary role may be to protect or sequester information required by the future germ cells. In addition, however, they could provide a convenient anchor for substances like the *nanos* mRNA that must be retained in the posterior region. Structures similar to the polar granules in *Drosophila* have been observed in association with germ plasm in many different invertebrate and vertebrate animals. Thus the *Drosophila* model could reveal basic concepts underlying the establishment of germline tissues in all species.

The National Institutes of Health provided financial support for two projects that are summarized here. These projects focus on the molecular and functional characterization of the *nanos* and *oskar* genes.

Localization of positional signals in the Drosophila *embryo. In situ hybridization experiment, using biotinylated cDNA probes, localizes RNA of the anterior signal* bicoid *and of the posterior signal* nanos *respectively to the anterior and posterior poles of an early (stage 2) embryo. Hybridization to* nanos *at the posterior also demarcates the position of the germ plasm.*

Research by Laura Dickinson. Photograph by Ruth Lehmann.

Regulation of Gene Expression During Cellular Differentiation and Activation

Jeffrey M. Leiden, M.D., Ph.D.—Associate Investigator
Dr. Leiden is also Associate Professor of Internal Medicine at the University of Michigan Medical School. He received his M.D. and Ph.D. degrees from the University of Chicago. His residency training was at Brigham and Women's Hospital, Boston, and his postdoctoral fellowship was in the laboratory of Jack Strominger at Harvard University.

THE processes of cellular differentiation and activation are accompanied by complex and precisely orchestrated changes in gene expression. Abnormalities in the patterns of expression of these specific genes may be involved in the etiology of a number of pathologic states, including autoimmune disease and malignancy. My laboratory is studying gene regulation during T lymphocyte and muscle cell differentiation in order to increase understanding of the molecular mechanisms that regulate gene expression during both normal and pathologic development.

Regulation of the T Cell Receptor Gene During T Cell Development

Human T lymphocytes recognize foreign antigens, such as virus-infected cells and tumor cells, via a specific cell-surface molecule, the T cell receptor (TCR). T cells can be divided into two subsets based on their expression of two distinct types of antigen receptor. The majority of T cells that circulate in peripheral blood, including all helper and cytotoxic T cells, express the α/β heterodimer receptor, while a small but distinct T cell subset of unknown function expresses the γ/δ receptor. The α/β and γ/δ T cells appear to develop as separate lineages during thymic ontogeny.

During the last several years, my laboratory has been interested in identifying and characterizing the molecular mechanisms that regulate the expression of these different TCR genes during T cell development. In an initial set of studies, we identified the transcriptional enhancer elements that control the expression of the TCR α and β genes. These two enhancers were shown to be required for the expression of the α and β genes and to function equally well in both mouse and human cells. The identification and localization of the human TCR enhancers led us to propose that certain T cell tumors that had previously been shown to contain chromosomal translocations into the human TCR α and β loci might be caused by the apposition of the TCR gene enhancers with translocated human proto-oncogenes.

Our more recent studies have focused on precisely identifying and characterizing the enhancer DNA sequences and the nuclear proteins they bind that are responsible for controlling TCR α and β gene expression. These experiments have demonstrated that both enhancers contain 4–5 different binding sites for nuclear proteins. At least two of these sites in each enhancer are absolutely required for enhancer activity. Several of the nuclear protein–binding sites in each enhancer correspond to previously defined enhancer motifs, while others represent novel sequence elements. At least one site in each enhancer was shown to bind T cell–specific nuclear proteins.

Both enhancers were shown to contain a binding site that was identical to the previously described cAMP response element (CRE). To learn more about the function of this element in regulating T cell gene expression, we have cloned two novel CRE-binding proteins, CREB-2 and CREB-3, that specifically bind to the TCR α CRE and to CREs from several other eukaryotic promoters. Both of these new CREB proteins contain similar basic DNA-binding domains and a leucine zipper region that allows them to form dimeric complexes both with themselves and, potentially, with other members of the CREB protein family.

During the past year, we have shown that *ets*-1, a previously described human proto-oncogene, is actually a DNA-binding protein that specifically recognizes one of the nuclear protein–binding sites in the TCR α enhancer. These studies have helped to define the Ets proteins as transcriptional regulators that may play an important role in controlling T cell gene expression. More recently, we have cloned several new members of the Ets family of transcription factors. One of these, which we call Elf-1, regulates a set of genes during the process of T cell activation. In addition, it appears to play an important role in regulating the expression of HIV-2, one of the AIDS viruses, in T cells. Finally, in collaboration with Stuart Orkin (HHMI, Children's Hospital, Harvard Medical School), we have identified a

novel T cell–specific transcription factor called hGATA-3 that binds to a third nuclear protein–binding site within the TCR α enhancer. Ongoing studies are designed to determine the role of each of these transcription factors in T cell tumors, as well as in normal T cell development and activation.

Genetically Engineered Myoblasts as a Recombinant Protein Delivery System

A variety of acquired and inherited diseases are currently treated by repeated intravenous or subcutaneous infusions of recombinant or purified proteins. In addition to hemophilia A, which is treated with intravenous infusions of human factor VIII, these include diabetes mellitus, treated with subcutaneous or intravenous injections of insulin, and pituitary dwarfism, treated with subcutaneous injections of growth hormone. The development of cellular transplantation systems that could stably produce and deliver such recombinant proteins into the systemic circulation would represent an important advance in our ability to treat such diseases.

The ideal recombinant protein delivery system would utilize a cell that could be easily isolated from the recipient, grown and transduced with recombinant genes *in vitro*, and conveniently reimplanted into the host organism. Such a cell should produce large amounts of secreted recombinant protein, and following secretion, this protein should gain access to the circulation. Finally, these implanted, genetically engineered cells should survive for long periods and continue to secrete the transduced protein product without themselves interfering with the function of the tissue into which they were implanted. Several different cellular systems have been used to produce recombinant proteins *in vivo*. These include keratinocytes, skin fibroblasts, hepatocytes, lymphocytes, and bone marrow. Although some of these systems have yielded detectable levels of circulating proteins briefly, stable physiological levels of circulating proteins have proved difficult to produce in normal animals.

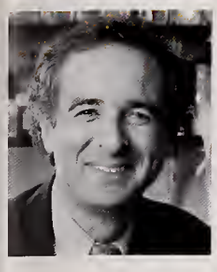
Genetically engineered myoblasts represent a potentially useful system for the *in vivo* delivery of recombinant proteins into the circulation. Myoblasts can be readily isolated from a muscle biopsy and expanded *in vitro* to very large cell numbers. Cultured myoblasts can be transfected *in vitro* and will synthesize large amounts of recombinant proteins. Most importantly, previous studies have demonstrated that cultured myoblasts can be injected intramuscularly and will survive and fuse into adjacent normal muscle fibers at the site of injection. Finally, skeletal muscle is a highly vascular tissue. Thus proteins secreted from myoblasts should readily enter the circulation.

We have recently explored the feasibility of using genetically engineered myoblasts as a recombinant protein delivery system. To this end, stable transfectants of the murine C2C12 myoblast cell line were produced that synthesize and secrete high levels of human growth hormone (hGH) *in vitro*. These stably transfected myoblasts were injected intramuscularly into normal syngeneic C3H mice, and serum and muscle levels of hGH were measured 5 days to 3 weeks after injection.

Mice injected with the growth hormone–transfected myoblasts produced significant and physiological levels of hGH both locally in muscle and in serum, as compared with control mice injected with nontransfected myoblasts. Human growth hormone levels in both muscle and serum were stable for at least three months following injection and exceeded those measured in serum from normal human volunteers. Histological examination of muscles injected with β-galactosidase–expressing C2C12 myoblasts demonstrated that many of the injected cells had fused to form multinucleated myotubes. Thus these studies demonstrated that genetically engineered myoblasts represent a novel and powerful system for the stable delivery of recombinant proteins into the circulation.

Dr. Leiden is now Professor of Medicine and Pathology at the University of Chicago.

Chemical Communication



Michael R. Lerner, M.D., Ph.D.—Associate Investigator
Dr. Lerner is also Associate Professor in the Departments of Internal Medicine and of Pharmacology and the Child Study Center at Yale University School of Medicine. He obtained his B.A. degree in chemistry from the University of Pennsylvania and his M.D. and Ph.D. degrees from Yale. His doctoral research, with Joan Steitz, was on small nuclear ribonucleoproteins (snRNPs). He interned in internal medicine at Barnes Hospital, St. Louis, and did postdoctoral research in neurobiology at Washington University with Gerald Fischbach before returning to Yale. His honors include the George Herbert Hitchings Award for innovative methods in drug design.

INTRASPECIES communication via specific chemical messengers is widely employed throughout the animal kingdom. Among the common uses of chemical communication are marking of territory, signaling danger, and indicating sources of food. A particularly striking example is the use of sex pheromones. Here, animals release a defined blend of related molecules that trigger distinct mating behaviors in members of the opposite sex. For detection to occur, three criteria must be satisfied. Molecules of the pheromone must reach olfactory receptors, must interact with them, and must be inactivated so that subsequent molecules can be detected.

Past research has focused on the problems of chemical transport to olfactory receptors and inactivation of these molecules. Moths—particularly *Manduca sexta* and *Antheraea polyphemus*—have provided excellent models. Many of the olfactory receptor cells of the male *M. sexta* and most of those from the male *A. polyphemus* are specialized for detecting sex pheromone. For both animals, the pheromone-binding proteins, which solubilize molecules of the pheromone blend and carry them to receptors, have been characterized. In addition, a family of general odorant-binding proteins, which are related to the ones employed by the moths to carry pheromone, has been discovered and characterized. Likewise, the enzymes that rapidly and specifically inactivate pheromone molecules and are apparently requisite to the sensory apparatus employed by males to locate females, have been investigated at the biochemical level. For *M. sexta*, in which both major components of the pheromone blend are aldehydes, a single enzyme, an aldehyde oxidase, suffices, while *A. polyphemus* employs both an aldehyde oxidase and an esterase because its pheromone components include both an aldehyde and an ester. Our research has now turned to developing methods for investigating olfactory receptors themselves.

Over the past few years many laboratories have conducted biochemical, electrophysiological, and molecular cloning experiments concerned with the nature of signal transduction by olfactory receptors. Depending on the animal whose sense of smell is being investigated, the results indicate that either cAMP, IP_3/DAG (IP_3, inositol 1,4,5-trisphosphate; DAG, diacylglycerol), or both second messenger systems are involved. It now appears that the study of olfaction, and hence a major aspect of chemical communication between animals, is part of the general problem of how G protein–coupled receptors work.

At the present time, several methods are employed to study how receptors that regulate intracellular concentrations of cAMP or IP_3/DAG work, such as adenylate cyclase assays, radioimmunoassays, measurements of IP_3 or DAG, monitoring the flow of ions through channels in frog oocyte membranes, or looking at changes in intracellular calcium concentrations via appropriate fluorescent indicators. To study G protein–coupled receptors, and ones relevant to olfaction in particular, we are developing a new method for monitoring receptor stimulation that can track changes in intracellular concentrations of cAMP or IP_3/DAG in over 10,000 individual cells simultaneously.

To follow changes in cAMP or IP_3/DAG levels in many individual cells at the same time, we have turned to that ability of some animals to change colors rapidly. In nature, color changes are used for such purposes as camouflage and the communication of states of emotion. Among vertebrates, quick color alterations are brought about by the controlled movement of pigment granules within chromatophores. When pigment granules in melanophores (a particular type of chromatophore) are aggregated, the animal appears light, and when pigment is dispersed, the animal appears dark. It turns out that the pigment translocation apparatus is controlled via second messenger systems that are themselves regulated by G proteins. As a result, the state of pigment disposition within melanophores reflects the state of activity of G protein–coupled receptors.

Recently the laboratory has successfully harnessed frog melanophores as the center of an

assay for monitoring the activity of G protein–coupled receptors that act to raise or lower intracellular cAMP or that raise intracellular IP_3/DAG. The ability of exogenous receptors to couple to and control the pigment translocation apparatus within melanophores should lead to improved methods for studying G protein–coupled receptors.

Structural Determinants of Human α-Globin Gene Expression

Stephen A. Liebhaber, M.D.—Investigator

Dr. Liebhaber is also Professor of Genetics and Medicine (Hematology) at the University of Pennsylvania School of Medicine. He received his B.A. degree in chemistry from Brandeis University and his M.D. degree from Yale University. He took clinical training in internal medicine and hematology at Case Western Reserve, the University of Colorado, Washington University, and the University of California, San Francisco. As a postdoctoral fellow with David Schlessinger at Washington University, Dr. Liebhaber examined ribosomal RNA processing, and with Yuet Wai Kan at the University of California, San Francisco, he studied human globin gene expression and genetic defects in α-thalassemia. Before moving to Philadelphia, he was a faculty member of the Department of Medicine at UCSF.

OUR laboratory has largely concentrated on studying the expression of the human globin genes. These genes encode hemoglobin, the major red cell protein responsible for transport of oxygen from the lungs to peripheral tissues. Since the hemoglobin molecule, $\alpha_2\beta_2$, is composed of an equal number of α- and β-globin chains, normal synthesis demands balanced expression of both sets of genes, which are located on different chromosomes. Defects in either set result in an imbalance of expression and consequent hereditary anemia: α- or β-thalassemia. Thalassemias result from more than 150 different mutations in the globin genes, affecting the health of millions worldwide.

Certain characteristics of globin gene expression make it particularly interesting for study. The extremely high level of globin mRNA in the differentiating red cell (over 95 percent of total cellular mRNA) has no equal in any other cell type. This abundance reflects both high levels of globin gene transcription and an unusual stability of the mature globin mRNA.

An additional interesting aspect of globin expression is that the genes in the α- and β-globin gene clusters follow an orderly sequence of expression during embryologic development. This results in a well-defined switch from embryonic to adult globin gene expression during development of the fetus. The switching results in the synthesis of successive hemoglobins with oxygen affinities that match changes in the uterine environment. The active transcription, the clearly defined pattern of developmental switching, and the unusual mRNA stability are areas of special focus in our laboratory.

The loss of α-globin expression observed in α-thalassemia usually results from deletion or abnormal structure of one or more of the α-globin genes. We have recently studied three unrelated patients from Germany, Portugal, and Italy who have an unusual form of α-thalassemia. In each case the loss of α-globin synthesis reflects loss of expression of a structurally normal α-globin cluster. In other words, one of the α-globin gene clusters in each of these patients is not functioning, even though the genes in these clusters are structurally normal and synthesize normal levels of α-globin when isolated and expressed in tissue culture cells. Although initially puzzling, the mechanism of this remarkable expression defect was eventually defined by extensive DNA mapping. In each of the three independent cases we demonstrated a large deletion 5′ to the inactivated cluster itself. In one case this deletion began as much as 50,000 bases 5′ to the silenced α-globin genes. By comparing the maps of each of these deletions, a region of common overlap was noted. These studies define a new category of α-thalassemia and demonstrate a critical determinant of α-globin gene expression located entirely external to the α-globin gene cluster. A similar set of transcriptional signals has been localized adjacent to the β-globin gene cluster. One can therefore speculate that such signals serve coordinately to activate and balance the expression of the α- and β-globin clusters in the red cell.

The human α-globin gene cluster contains a ζ-globin gene expressed specifically in the embryo and two α-globin genes, α1 and α2, expressed in the fetus and the adult. The switch from embryonic ζ-globin to adult α-globin expression occurs at the end of embryonic development (7–8 weeks of gestation). This critical developmental switch, which occurs widely in mammals, presents a well-defined model system for studying developmental control of gene expression.

To establish a system in which to study switching within the human α-globin gene cluster, we have injected the human embryonic ζ- and adult α-globin genes into fertilized mouse eggs to generate transgenic mice. The red cells of these transgenic mice appropriately express the human transgenes during development. In the embryonic period, there is a parallel expression of the human and mouse ζ-globin genes and, by day 12 of development, parallel expression of the α-globin genes. These data suggest that 1) the hu-

man transgenes contain the necessary information for appropriate developmental control and 2) the factors responsible for developmental switching in mouse red cells have been sufficiently conserved during evolution to substitute in the control over the human transgenes. By generating transgenic mice that carry only the human α-globin gene in the absence of the ζ-globin gene, we have further demonstrated that the appropriate expression of this gene is not dependent on competition between the two developmentally specific genes. By introducing into the mouse genome more limited gene fragments, as well as genes with specific alterations, and by studying their pattern of developmental expression in the transgenic model system, we should be able to define the signal(s) critical to globin gene switching.

Selective stabilization of globin mRNA is essential to its accumulation in terminally differentiating red cells to 95 percent of total cellular mRNA. We are attempting to define the structural basis for the stability of α-globin mRNA through investigation of an α-thalassemia mutation in which this stability is lost: α Constant Spring (αCS). This mutation, the most common cause of nondeletional α-thalassemia in Southeast Asia, is a CAA for UAA substitution at the normal termination codon of the α2-globin gene. As a result of this single-base substitution, the ribosome translates into the normally untranslated 3′ region and destabilizes the mRNA.

To study the basis for this destabilization in detail, we have established an experimental system that reproduces the selective instability of αCS mRNA in tissue culture cells. Remarkably, we find that the instability of the αCS mRNA relative to normal α-globin mRNA observed in the patient is recapitulated faithfully when the α- and αCS-globin genes are expressed in an erythroid tissue culture cell line. In contrast, we find that the mRNAs from these two genes are expressed at equivalent levels when expressed in nonerythroid cell lines. These data suggest that the stability of globin mRNA may depend on interaction with one or more erythroid-specific factors.

By specifically altering the structure of the α-globin genes prior to expression, we have demonstrated that the region critical to the stability of the α-globin mRNA is located in a segment of the 3′-nontranslated region just past the translation termination codon. This determinant can be destabilized by the translating ribosome if it is allowed to enter this region, as is the case in the αCS mutation, or by site-specific alterations in the primary sequence of this segment. These data suggest that it should now be possible to map the limits of this determinant and subsequently identify critical cellular factors mediating this response.

The Heat-Shock Response

Susan L. Lindquist, Ph.D.—Investigator

Dr. Lindquist is also Professor in the Department of Molecular Genetics and Cell Biology and in the Committees on Developmental Biology and Genetics at the University of Chicago. She received her B.S. degree in microbiology from the University of Illinois, where she worked with John Drake on bacteriophage T4. Her graduate research was done with Matthew Meselson at Harvard University, where she began her studies on the heat-shock response. She continued this work during her postdoctoral research with Hewson Swift at the University of Chicago.

THE causes of heat-induced lethality and the mechanisms that cells employ to protect themselves from heat damage are poorly understood. Over the past decade, a great deal of research has focused on a small group of molecules, the heat-shock proteins (HSPs). These are induced in response to temperature elevation and a wide variety of other stresses. No known genetic induction is more highly conserved in evolution, which underscores its fundamental importance in biology. Archaeobacteria, eubacteria, plants, and animals all produce similar proteins. Several of these proteins show very high levels of conservation, commonly with 40–50 percent amino acid identity between HSPs of human and bacterial cells.

The induction of HSPs 1) allows cells to grow at the upper end of their normal growth range, 2) potentiates survival during long exposures to temperatures just beyond the normal growth range, and 3) protects cells from lethality at temperature extremes. Interestingly, different proteins are required for each of these functions. HSPs also protect cells from heavy metal ions, ethanol, and many other sources of stress. The importance of different proteins in protecting against different forms of stress also varies.

The stress inductions of HSPs are of interest to human biology and medicine for several reasons. First, studies of cultured cells *in vitro* and of tumors *in vivo* demonstrate that many cancer cells are more readily killed by heat than are untransformed cells. For this reason, hyperthermia, in conjunction with radiation and chemotherapy, is emerging as an important new tool in cancer therapy. Second, high temperatures are associated with a number of developmental anomalies in a wide variety of plants and animals, including spina bifida in humans. In those organisms that have been subjected to experimental manipulation, mild preheat treatments, which induce the HSPs, provide protection. Third, the induction of HSPs is associated with a variety of human pathological states, including strokes, heart attacks, and kidney disease. Interest in the proteins includes both their putative protective functions in affected tissues and the possibility of quantifying them as disease markers. Fourth, the proteins interact with and potentiate the function of many other vital proteins in the cell.

The heat-shock response also provides a superb model system in which to study the cellular mechanisms involved in regulating protein synthesis. Because HSPs are required for survival, a number of regulatory mechanisms are employed to ensure that the proteins will be produced as rapidly as possible after exposure to stress. Thus studies of the response have provided fundamental insights on the nature of nuclear and cytoplasmic regulation in both eukaryotes and prokaryotes.

The recent discovery that the HSPs themselves, or close relatives produced at normal temperatures, play vital roles during normal growth and development has opened up a whole new field of investigation. The specific molecular functions of the HSPs are only beginning to be elucidated, but they play a role in a remarkable number of basic cellular processes, including secretion, signal transduction, and ribosome assembly. Determining the roles that HSPs play in these processes will provide fundamental insights in cell biology.

We are investigating the regulation and function of these proteins. Our research focuses on the yeast *Saccharomyces cerevisiae* and the fruit fly *Drosophila melanogaster,* because techniques of genetic manipulation and molecular analysis are so advanced in these organisms. For the past few years our investigations of the regulation of the response have concentrated on post-transcriptional mechanisms that are employed to maximize the synthesis of the HSPs during heat shock or to shut off synthesis after heat shock.

Tom McGarry and Bob Petersen found that heat-shock mRNAs in *Drosophila* cells are preferentially translated during heat shock by virtue of sequences in their 5′-untranslated leaders and are preferentially repressed during recovery through sequences in their 3′-untranslated tails. The latter sequences are shared by certain normal cellular messages, which have the common property of being rapidly degraded at normal temperatures.

Heat-shock regulation takes advantage of this common pathway to control HSP expression. The mechanism is inactivated during heat shock and restored during recovery.

Joseph Yost demonstrated that heat shock blocks the processing of mRNA precursors, which explains why heat-shock genes generally do not have intervening sequences. (If they had, the block in splicing would prevent expression.) Sudden high-temperature heat shocks also inhibit transcription termination (discovered by Robert Dellavalle). By some mechanism we do not yet understand, heat-shock genes are more likely to be terminated correctly than normal cellular genes.

Although our studies of HSP function are in many ways independent of our studies on regulation, in one important respect they overlap. HSPs, and hsp70 in particular, play an important role in restoring normal gene expression patterns after heat shock. They are required at the level of translation, RNA turnover, RNA processing, and transcription.

To investigate the function of the HSPs, we created a series of mutations in the genes of both yeast and *Drosophila*. Kathy Borkovich found that hsp82 is essential for growth at all temperatures in yeast cells, but is required at higher concentrations for growth at high temperatures. Thus induction is required for cells to grow at the upper end of their normal temperature range. We believe the protein is needed to regulate the activity of a wide variety of other cellular proteins and that it is needed at higher concentrations at high temperatures in order to drive the equilibrium of these interactions toward complex formation.

In collaboration with Keith Yamamoto's laboratory, Bushra Khursheed and Marc Fortin demonstrated that hsp82 interacts with the steroid hormone family of receptors and helps these proteins fold into an active conformation. Most recently, Yang Xu found that hsp82 is also required for the maturation of oncogenic proteins in the src family.

Yolanda Sanchez created mutations in the *HSP104* gene of yeast. The mutations have no effect on growth at high or low temperatures. However, the cells are unable to acquire tolerance to extreme temperatures when given a mild preheat treatment. Thus this mutation confirms the long-standing assumption that HSPs play a vital role in establishing thermotolerance. Moreover, it provides protection from many other forms of stress, such as exposure to ethanol and sodium arsenite. Dawn Parsell found that hsp104 is highly conserved in mammals and in prokaryotic cells and contains two essential nucleotide-binding domains.

In *Drosophila* our mutational analysis has concentrated on hsp70. Janice Rossi found that varying the level of hsp70 expression in *Drosophila* cells varies the rate at which the cells recover from heat shock. Jonathan Solomon found that expressing hsp70 from independently regulated promoters, in the absence of heat shock, helps cells to survive extreme temperatures but inhibits their growth. Thus hsp70 helps to protect cells from the ravages of extreme temperatures but is actually disadvantageous at normal temperatures.

To study the role of hsp70 in whole flies, Kent Golic developed a new system for manipulating the *Drosophila* genome. He created flies that express the *FLP* recombinase gene of yeast under the control of heat-shock regulatory sequences. When flies that also carry a recombinase target sequence are given a very mild heat shock, the recombinase is induced and catalyzes rearrangements of the target sequence.

Michael Welte and Joan Tetrault have employed this site-specific recombination system to study the role of hsp70 in induced thermotolerance. Strains that carry several extra copies of the hsp70 gene were constructed. Embryos from these lines survive heat treatments much better than wild-type embryos. This suggests that it will be possible to alter the stress tolerance of even developmentally complex species.

T Cell Surface Glycoproteins in Development and Viral Infections

Dan R. Littman, M.D., Ph.D.—Associate Investigator

Dr. Littman is also Associate Professor of Microbiology and Immunology and of Biochemistry and Biophysics at the University of California, San Francisco. Following undergraduate studies on the structure of microtubules in Marc Kirschner's laboratory at Princeton University, he completed his M.D./Ph.D. program at Washington University in St. Louis, working with Benjamin Schwartz and Susan Cullen on the function of histocompatibility molecules in antigen presentation. His postdoctoral research was done in Richard Axel's laboratory at Columbia University, where he isolated the genes for CD4 and CD8.

THE shaping of a mature repertoire of T lymphocytes capable of responding to pathogenic microorganisms involves a complex process of differentiation within the thymus. In this process, self-reactive cells are eliminated and cells that can react to foreign antigen complexed to molecules of the major histocompatibility complex (MHC) are selected to survive and migrate to peripheral lymphoid organs.

The signals involved in the different phases of thymic selection require the interaction of several thymocyte surface proteins with components of the thymic microenvironment. These cell surface molecules include the clonally restricted T cell receptors and the CD4 and CD8 glycoproteins. Both CD4 and CD8 are expressed on immature thymocytes, but the gene for one or the other is shut off upon maturation. Cells that have receptors for self-MHC class I molecules continue to express CD8, but shut off CD4; cells with receptors for MHC class II molecules express CD4 and shut off CD8. Our laboratory is studying the mechanism of regulation of the CD4 and CD8 genes as well as the molecular basis of the cell-cell interaction resulting in the selection of cells that have appropriate specificity.

The CD4 and CD8 molecules interact directly with MHC class II and class I molecules, respectively. Effective T cell activation requires corecognition of MHC by both the T cell receptor and either CD4 or CD8. To test whether corecognition is also required in transmembrane signaling during T cell development, we have analyzed transgenic mice that express a mutant form of MHC class I incapable of binding to CD8 but interactive with T cell receptors. These mice were shown to be defective in developing a mature T cell repertoire specific for the mutant MHC class I molecule.

This study indicates that CD8:MHC binding is required for both intrathymic deletion of self-reactive T cells and positive selection of useful T cells. Moreover, since endogenous MHC class I molecules were competent to bind CD8 but unable to rescue the defect, coordinate recognition of antigen/MHC by a complex of the T cell receptor and CD8 is required for both positive and negative selection.

To study the roles of CD4 and CD8 in greater detail, we are preparing animals defective in the expression of these molecules. We have utilized gene-targeting technology to mutate the CD4 gene in embryonal stem cells. These have been injected into mouse blastocysts, resulting in the birth of chimeric animals. The mutant CD4 gene has been propagated in the mouse germline, and we are currently studying the immune system of mice lacking CD4 expression. These mice lack helper T cells and have essentially no antibody response to T-dependent antigens and no T cell response to antigen presented by MHC class II molecules.

Transgenic mice expressing mutant forms of CD4, predicted to be defective in signal-transducing functions (as a result of defective interaction with cytoplasmic tyrosine kinases or with the T cell receptor complex), have been prepared and are being used to analyze the role of CD4 signaling during development in the absence of endogenous CD4 expression. We have also prepared transgenic mice that express the human CD4 gene and will examine the ability of the human molecule to function during T cell development in mice that lack murine CD4. Such animals may be especially useful for studies of pathogenesis and therapy in autoimmune diseases and AIDS.

The CD4 and CD8 glycoproteins have been shown to be involved in the activation of peripheral T lymphocytes. For example, artificial cross-linking of CD4 or CD8 to the T cell receptor complex results in T cell activation. Moreover, T cells that have lost expression of either CD4 or CD8 but continue to express the T cell receptor are usually no longer responsive to antigen. We are performing a variety of structure/function studies to determine the mechanism through which CD4 and CD8 facilitate signal transduction in T cells.

It is known that CD4 and CD8 are associated

with a cytoplasmic tyrosine kinase, $p56^{lck}$, a member of the *src* kinase family. The sites of interaction between these molecules map to cysteine-containing regions in the cytoplasmic domains of CD4 and CD8 and in the unique amino-terminal domain of $p56^{lck}$. We have demonstrated that only CD4 molecules that can associate with the kinase are functional in antigen-specific T cell hybridomas, which normally produce interleukin-2 upon stimulation with the appropriate antigen.

Since interaction of CD4 with the *lck* kinase is essential for T cell activation, we are now beginning to characterize the proteins that serve as substrates for tyrosine phosphorylation. One of the substrates appears to be a tyrosine kinase associated with the T cell receptor complex. Cross-linking of $p56^{lck}$ to the receptor-associated kinase results in tyrosine phosphorylation and activation of phospholipase-Cγ1, an enzyme that is also associated with the T cell receptor complex. Cleavage of phosphoinositol-containing phospholipid by this activated enzyme generates second messengers that activate the transcriptional machinery in the T cell.

The laboratory is also studying the mechanism involved in the developmental switch from double-positive ($CD4^+CD8^+$) to single-positive thymocytes. The switch may be due to differential signaling through CD4 or CD8 (hence due to instruction) or to random (stochastic) shutoff of either CD4 or CD8, followed by selection of cells that have the appropriate T cell receptor and co-receptor. Several experiments are being performed to discriminate between these two basic mechanisms of differentiation.

An understanding of the mechanism involved should also shed light on the transcriptional regulation that thymocytes employ, first to turn on both the CD4 and CD8 genes and subsequently to shut off only one of the genes. We have characterized a sequence upstream from the CD4 promoter with T cell–specific transcriptional enhancing activity. The ability of this and other sequences to provide subset-specific regulatory signals in thymocyte differentiation is being studied in transgenic mice.

The CD4 glycoprotein is doubly important because it is the receptor for the human immunodeficiency virus (HIV). We have found that several cell types that express CD4 can bind virus but cannot be infected, indicating that host factors other than CD4 are involved in viral entry. Resistance of these cells to infection is due to the inability of the viral envelope to fuse to the cellular plasma membrane. It is therefore likely that plasma membrane molecules other than CD4 are required for fusion of virus to target cells. We are using genetic and biochemical approaches to identify such molecules.

Identification of additional molecules involved in HIV entry may permit design of novel agents to interfere with the spread of HIV. In addition, we hope that this information will facilitate the design of a mouse model system for HIV disease. The currently available mice that express the human CD4 transgene are resistant to infection, but expression of additional genes involved in HIV entry may permit infection of these animals with HIV.

Using a genetic system for studying HIV entry, we have shown that the envelope glycoprotein of HIV can be replaced by that of another pathogenic human retrovirus, human T cell leukemia virus (HTLV), forming HIV(HTLV) pseudotypes. HTLV causes T cell leukemias and lymphomas and myelopathies. Individuals infected with both HIV and HTLV have more rapid progression of HIV disease than those infected with HIV alone. Since both viruses infect T lymphocytes, it is likely that mixed viral particles can form *in vivo*. We have demonstrated that mixed particles (pseudotypes) that form *in vitro* have an expanded host range—i.e., HIV particles, endowed with the HTLV-I envelope glycoprotein, can readily infect CD4-deficient cells. Such mixed particles may have an important role in HIV pathogenesis, particularly in infection of cells that do not have HIV receptors, such as cells of the central nervous system. In addition to investigating HIV infection, we are using the hybrid particles to study the yet uncharacterized HTLV receptor.

The Biology of T Lymphocyte Development

Dennis Y.-D. Loh, M.D.—Associate Investigator

Dr. Loh is also Professor of Medicine, Genetics, and Molecular Microbiology at Washington University School of Medicine and Chief of the Division of Allergy and Immunology and Associate Physician at Barnes Hospital, St. Louis. He received his undergraduate degree in biology and chemical engineering from the California Institute of Technology and his medical degree from Harvard Medical School. After finishing his clinical residency in internal medicine at the Peter Bent Brigham Hospital, Boston, he studied as a postdoctoral fellow with David Baltimore at the Massachusetts Institute of Technology.

THE immune system is involved intimately in our body's defense against invading microorganisms and tumors. In addition, it plays a central role in organ graft rejection and autoimmune diseases such as systemic lupus erythematosus, rheumatoid arthritis, and diabetes. Its critical role in maintaining health is best manifested in the acquired immune deficiency syndrome (AIDS), in which destruction of a specific portion of the immune system results in a potentially fatal disease.

Our investigation is focused on the molecular mechanisms that allow normal and abnormal development of the antigen-specific T cells. T cells are those lymphocytes (a type of white cell) that depend on the presence of the thymus gland for maturation. An antigen may be viewed as any marker that these cells recognize. T cells are thought to play a central role in the regulation of the immune response. T cells recognize antigens by means of a cell surface structure called the T cell receptor (TCR). The genes that are responsible for the expression of the TCR undergo DNA gene rearrangement and gene activation specifically in the thymus during the individual's early development. Once the TCR is expressed as protein, it is the interaction of the TCR with its antigen that triggers the activation of T cells, resulting in an immune response. The ultimate result of such a response may be either defense against invading organisms or tissue destruction, as seen in transplantation rejection and autoimmune phenomena.

My laboratory initially concentrated on identifying the genetic elements that encode the TCR genes. We then shifted our efforts to study the function of T cells in the intact animal, especially during its development. Two important questions were addressed. 1) Why are we tolerant of our own tissues and organs? 2) Why are transplanted organs rejected readily (unless they are carefully cross-matched)? We have used both recombinant DNA technology and our ability to create transgenic mice (mice with cloned genes incorporated in their own chromosomes) to study these questions. Two kinds of transgenic mice have been created. One kind bears transgenic TCR genes; the other has transgenic major histocompatibility complex (MHC) genes (a marker that distinguishes us individually during transplantation). By introducing these genes back into the mouse itself and into the mouse germline, we can determine how normal T cells develop by studying how the TCR and MHC interact during development.

This strategy has been very successful. By creating mice of appropriate genetic background, we discovered that T cells that are self-recognizing and hence self-reactive are deleted in the thymus during development. This implies that part of self-tolerance is accomplished by physical elimination of self-reactive T cells. Using mice containing transgenic MHC, we showed a second mechanism of self-tolerance that does not involve physical deletion. In this case, self-reactive cells are not physically eliminated but are functionally paralyzed. These studies allow us to lay the foundation to study how T cells acquire self-tolerance. Since distinguishing what is self and nonself is a central problem in immunology, we hope that these studies will lead to a better understanding of transplantation rejection and autoimmune phenomena.

The detailed study of T cell development also allows us to investigate the cellular and molecular requirements of normal cellular developmental processes. For example, we still do not understand how certain T cells are eliminated while others are selected to survive. To elucidate the exact molecular mechanisms that underlie this highly regulated process, we have most recently focused on the structural and signaling basis that controls differential cell fate of the developing thymocytes. To do this we have combined the two powerful technologies of transgenic mice and gene knock-out mice. In the latter mice, selected genes can be targeted to be destroyed in the germline so that mice bearing selected mutations can be created. Once these mice have been prepared, mutant molecules can be introduced to replace the "knocked-out" genes. Current efforts are concentrated on signaling molecules such as CD4 and CD8 that are important in T cell devel-

opment. By such an approach, we hope to understand the signaling requirements that determine cell fate during development as well as the exact mechanisms by which T cells undergo the programmed cell death that leads eventually to self-tolerance.

Molecular Genetics of Mammalian Glycosyltransferases

John B. Lowe, M.D.—Assistant Investigator

Dr. Lowe is also Associate Professor of Pathology at the University of Michigan Medical School. He received his bachelor's degree in mathematics from the University of Wyoming and his M.D. degree from the University of Utah College of Medicine, Salt Lake City. He was trained in clinical pathology and molecular genetics at Washington University School of Medicine, St. Louis. He was later Assistant Professor in the Departments of Pathology and Medicine at Washington University and also served as Assistant Medical Director of the Barnes Hospital blood bank, St. Louis, before moving to Michigan.

THE primary long-range goal of our research is to understand the functions of oligosaccharides that are found on the surface of mammalian cells and to explain how the cells regulate their expression. Oligosaccharide molecules consist of many different single-sugar structures linked together in complex linear and branching arrays. Quantitative and structural changes in such molecules have been shown to correlate with morphologic changes that occur during the embryonic development of animals and in association with neoplastic transformation. These and other observations suggest that cell surface oligosaccharides may function as information bearers in mediating interactions between cells during the developmental process.

Mammalian cells, in constructing these molecules, use special proteins called glycosyltransferase enzymes. With few exceptions, a unique glycosyltransferase is responsible for the synthesis of each linkage between the sugar molecules in an oligosaccharide. The enormous number of different oligosaccharides dictates that many different glycosyltransferases will enter the construction of the complex cell surface carbohydrates on any particular cell or tissue.

In many instances, changes in cell surface carbohydrate structure observed during differentiation or in association with malignant transformation have been shown to correlate with changes in the glycosyltransferase repertoire. The mechanisms by which cells coordinate and regulate the expression of these enzymes, and thus the expression of oligosaccharide structures at the cell surface, are largely unknown. During the past few years, the main focus of our work has been in establishing systems that will allow molecular analysis of the mammalian genes responsible for glycosyltransferase synthesis.

The human ABO, H, and Lewis blood group antigens are actually cell surface oligosaccharides. The determinant genes encode particular glycosyltransferases that are able to construct the "blood group" molecules. These glycosyltransferases provide convenient genetic and biochemical models for studying how the processes regulate cell surface oligosaccharide expression. The blood group antigens are not restricted in their expression to blood cells. They are found on a number of other tissues in the body, suggesting that tissue-specific mechanisms regulate their expression. Moreover, their expression changes during human embryonic development and is often altered in malignancy.

Our initial efforts focused on developing systems to isolate glycosyltransferase genes without the benefit of purified enzyme protein. Using gene transfer approaches, we have been able to isolate several of these genes. They include human genes encoding the H blood group $\alpha(1,2)$fucosyltransferase and the Lewis blood group $\alpha(1,3/1,4)$fucosyltransferase.

The cloned gene segments in each case represent tools for investigating the genetics of these enzymes and for studying the function and regulation of their corresponding cell surface oligosaccharides. For example, we have recently used the H blood group gene to investigate the molecular basis for the Bombay blood group phenotype. Individuals with this blood group are extraordinarily rare and are cross-match incompatible with virtually all other humans, excepting other Bombay individuals. This incompatibility is due to the fact that these persons apparently lack a functional H blood group locus. As a consequence, they are unable to construct A, B, or H blood group determinants, and thus maintain high titers of antibodies directed against the ABH blood group structures found on red cells from virtually all other humans.

The molecular basis for the defect in Bombay individuals had not been defined. By analyzing the structure of the H gene in Bombay pedigrees, we identified point mutations in both alleles of the gene in affected individuals. We subsequently demonstrated that these mutations inactivate the enzyme encoded by the gene and are thus responsible for the Bombay phenotype. We have also analyzed this gene in para-Bombay individuals, whose red cells are deficient in ABH structures but whose secretory tissues express essentially normal levels of these molecules (under

the control of the Secretor blood group locus). We also found inactivating point mutations in the H gene in these persons, thus providing strong evidence that the human Secretor blood group locus corresponds to a distinct $\alpha(1,2)$fucosyltransferase gene.

We have also used glycosyltransferase gene segments to identify specific cell surface oligosaccharide molecules that play pivotal functional roles in the inflammatory process. One of the primary events in inflammatory conditions involves a process whereby circulating white cells leave the interior of blood vessels and become localized in inflammatory foci outside the vascular system. This process begins when the endothelial cells lining the blood vessels become "activated" by substances that accompany an incipient inflammatory condition. Circulating white cells known as neutrophils adhere tightly to activated endothelium, insinuate themselves into the endothelial cell pavement lining the blood vessel, and ultimately come to occupy areas outside the vascular tree.

Neutrophil adhesion to activated endothelium is mediated in part by a protein known as E-selectin, or endothelial leukocyte adhesion molecule 1 (ELAM-1), which is found on the surface of activated endothelium. Structural features exhibited by this protein suggested that it might interact with an oligosaccharide molecule specific to the surface of neutrophils. By transfecting different glycosyltransferase gene segments into mammalian host cells, we were able to recapitulate the biosynthesis of several distinct sets of cell surface oligosaccharide molecules and to demonstrate that one set allowed transfected cells to adhere to E-selectin. We further demonstrated that the oligosaccharide molecules were one or more members of a family of oligosaccharides containing sialic acid and fucose and were represented by a molecule known as the sialyl Lewis X determinant. More recently, we have shown that E-selectin maintains a relatively high specificity for oligosaccharide ligands with $\alpha(1,3)$-linked fucose residues in terminal positions, versus subterminal locations. We have also shown that terminal $\alpha(1,3)$fucosylation may be catalyzed by some, but not all, human $\alpha(1,3)$-fucosyltransferases.

Aberrant or overexuberant recruitment of neutrophils to sites of inflammation can contribute, after tissue hypoxia and in other pathological circumstances, to undesirable tissue damage in autoimmune disease. Initial events in this process require adhesive interactions between E-selectin and its oligosaccharide counter-receptors on the surfaces of neutrophils. It thus seemed possible that purified carbohydrate counter-receptors might function as anti-inflammatory molecules by preventing pathologic neutrophil recruitment. We have tested this notion in collaboration with Peter Ward here at the University of Michigan. We have prepared therapeutic quantities of sialyl Lewis X–containing oligosaccharides, using recombinant $\alpha(1,3)$fucosyltransferases. When tested in an animal model of E-selectin–dependent lung inflammation, these carbohydrates exhibit potent anti-inflammatory activity. This work suggests that a new class of antiinflammatory pharmaceuticals may be developed from such molecules or from their chemical analogues.

These studies have been aided by our recent identification of several novel human $\alpha(1,3)$fucosyltransferase genes. We have shown that these encode enzymes with shared and unique primary sequence domains, as well as distinctive and useful catalytic properties. For example, one of these works extremely well for *in vitro* synthesis of the sialyl Lewis X tetrasaccharide, and others can efficiently construct analogues of this molecule. Work is in progress to identify sequence domains within the enzymes that dictate their distinctive substrate specificities.

Circumstantial evidence gathered by other investigators suggests that oligosaccharides are importantly involved in cell adhesion during mammalian embryogenesis. We are now directing our efforts toward exploring this hypothesis and characterizing the genes that determine these interactions, through genetic manipulation of the murine genome.

Mechanisms of Embryonic Induction in Vertebrates

Richard L. Maas, M.D., Ph.D.—Assistant Investigator

Dr. Maas is also Assistant Professor of Medicine at Harvard Medical School and Associate Physician at Brigham and Women's Hospital, Boston. He received his A.B. degree in chemistry from Dartmouth College and an M.D.-Ph.D. degree from Vanderbilt University School of Medicine. Following his thesis work with John Oates, he trained as a medical house officer at Brigham and Women's Hospital and completed a postdoctoral fellowship in Philip Leder's laboratory in the Department of Genetics at Harvard Medical School.

THE goal of our research is to understand the role that the homeobox genes play in controlling vertebrate organogenesis. These genes are defined by their expression of a 60–amino acid, helix-turn-helix, DNA-binding domain. Highly conserved in evolution, they are present in species as divergent as *Drosophila*, yeast, and humans. Mutations in homeobox genes of fruit fly and mouse result in specific developmental defects. Our work thus far has focused on the characterization of two such genes that appear to play important roles in the formation of the mammalian kidney and eye, respectively. A long-term goal is to understand the target genes that these homeobox genes interact with, using a combination of biochemical, embryologic, and genetic techniques.

Murine Homeobox Genes Expressed in Mouse Embryonic Kidney

To determine which homeobox genes are expressed in the developing mouse kidney, we undertook a polymerase chain reaction (PCR) screen of reverse-transcribed, microdissected kidney RNA from mouse embryos at day 15. This experiment yielded 27 different homeobox-containing genes, some 11 of which correspond to new genes. Among these novel sequences, we identified several with 85–98 percent sequence similarity to known murine *Hox* genes at the nucleotide level. In addition, we identified two other genes, closely related to each other, that appear to be new members of the *Hox-1* and *Hox-3* clusters in the mouse. Mapping experiments indicate that the *Hox-1* member is *Hox-1.8*.

The structures of several cDNA clones of *Hox-1.8* are being determined. Thus far, five different alternatively processed forms have been identified. Surprisingly, all these forms share a common feature: due to the presence of upstream termination codons, none would actually encode a translatable homeodomain. We suspect that a homeodomain-encoding form exists, because the homeodomain is preserved intact at the sequence level. Current efforts are aimed at securing the 5′ end of the *Hox-1.8* gene in order to determine whether splice forms exist that would encode a functional homeodomain.

A current working hypothesis of our laboratory is that many murine *Hox* genes may, as a general rule, encode both homeobox-containing and homeobox-less forms. Such forms may interact in heterodimeric combinations with one another, or with other *Hox* genes, to affect the capacity of the homeobox-containing form to bind to DNA.

The expression of the *Hox-1.8* transcripts has been analyzed as a function of mouse embryogenesis. As determined by Northern blot analysis, significant expression appears at day 10 of embryogenesis, peaks at day 13, and subsides by day 15. Expression in adults is confined to the kidney and to skeletal muscle. Interestingly, the adult kidney appears to express a smaller transcript form of approximately 1.5 kb, in addition to the larger class of 2.8–3.9 kb observed in both embryos and kidney. This size range corresponds to extensive alternate processing, as noted above.

We have further analyzed the expression of *Hox-1.8* during embryogenesis by *in situ* hybridization. Regionally restricted expression is observed in the submucosa of the foregut and midgut and also in somites. Of particular interest to the potential role of this gene in nephrogenesis is its expression in the condensing collecting duct system, in the region that comprises the developing calyces. Three-dimensional reconstruction of the *Hox-1.8* expression pattern shows that expression is localized to this part of the developing kidney at embryonic day 13.

Identification of a *Pax* Gene Involved in Formation of the Vertebrate Eye

The formation of the vertebrate eye has long served as an attractive model system for studying basic features of embryonic induction. The eye forms as a consequence of outgrowth of the diencephalon and a subsequent interaction of this neuroectoderm-derived structure with the surface ectoderm, resulting in an invagination of the latter to form the lens vesicle. Additional neuroectoderm and mesodermal ingrowth anterior to the

lens results in formation of the anterior structures of the eye, specifically the iris, cornea, and ciliary body. An additional interaction, between the developing lens and retina, is suggested by experiments involving transgenic mice: ablation of the developing lens results in an abnormal proliferation of retina, perhaps suggesting the presence of an inhibitory factor from lens.

In both mouse and human, there are naturally occurring mutations that affect eye development. In the mouse, a mutation on chromosome 2, called *Small eye* (*Sey*), results in deficient eye formation. The portion of mouse chromosome 2 to which *Sey* maps is homologous to a portion of human chromosome 11, 11p13. Interestingly, a semidominant human disorder called aniridia, which also affects basic eye development, is located in this region of 11. The observation that a mouse paired box and homeobox gene, *Pax-6*, maps to the *Sey* locus prompted us to examine the role of this gene and its human counterpart in basic ocular development. Recently it has been shown that point mutations in the mouse *Pax-6* gene are in fact responsible for the *Sey* mutation.

We have cloned the human *PAX6* gene, which maps to within 200 kb or less of the human aniridia locus, as defined by various translocation breakpoints. This gene is more than 95 percent conserved over 422 amino acids with a related gene cloned from zebra fish, which is expressed in both the diencephalon and lens vesicle during ocular development. Thus both the chromosomal location and the expression pattern of this gene are consistent with its involvement in aniridia.

We have determined the genomic structure of the *PAX6* gene, which consists of at least 13 exons spread over some 30 kb of genomic DNA, and have sequenced the intron-exon boundaries. This has permitted the design of PCR primers to each exon, which are conveniently sized for analysis by PCR. Thus far we have identified a nonsense mutation located in helix 2 of the homeobox of the human *PAX6* gene in one patient and, preliminarily, a small insertion in the *PAX6* gene of another patient. Although analysis of more mutations will be required to establish that mutations in *PAX6* indeed account for aniridia, these findings argue persuasively that they do.

Cell Cycle Control

James L. Maller, Ph.D.—Investigator

Dr. Maller is also Professor of Pharmacology at the University of Colorado School of Medicine. He received his B.S. degree in biochemistry from Cornell University and his Ph.D. degree in molecular biology from the University of California, Berkeley, where he worked with John Gerhart. He then carried out postdoctoral studies with Edwin Krebs at both the University of California, Davis, and the University of Washington before moving to Colorado.

TWO events mark the reproductive life of a cell: replication of the DNA, and its distribution to daughter cells at mitosis. Because of the central importance of cell reproduction to ordered cell growth and to birth of the next generation, cells have evolved rigorous controls to ensure that both events are carried out with high fidelity and at the appropriate time. My laboratory is interested in understanding the nature and regulation of these controls with respect to how a cell commits itself to replicate its DNA and how it knows when to divide.

The cell cycle has four main phases: G_1, S (synthesis), G_2, and mitosis. The decision to synthesize DNA (to enter the S phase) is made in G_1, and the decision to begin cell division (to enter the M phase of mitosis) is made in G_2. There is abundant evidence that these decisions are made at checkpoints, or restriction points, in the cycle. The nature of these G_1 and G_2 decision-making periods in the cell cycle underlies fundamental processes operative in early embryonic development and in malignant cells.

$G_2 \rightarrow$ M Regulation

Our laboratory developed a G_2-phase extract from frog eggs in which synthetic nuclei entered mitosis at the addition of mitotic signals. We then purified the mitosis-signaling enzyme (called maturation-promoting factor, or MPF) and found that it was composed of a protein kinase complexed to a G_2 cyclin. Kinases have the ability to attach a phosphate group to many different cellular proteins, modifying their function and causing profound changes in cellular biochemistry. The protein kinase was identified as a vertebrate homologue of the *cdc2* gene, which had been genetically implicated in the control of mitosis by the study of certain mutants in yeast.

G_2 cyclins are proteins that accumulate during interphase, reach high levels in late G_2, and are then degraded near the metaphase $\rightarrow$ anaphase transition in mitosis. This degradation is required in order for cells to complete mitosis successfully and enter G_1. In most cells there are two classes of G_2 cyclins, termed A and B cyclins, that differ in sequence similarity and have different kinetics of accumulation and degradation. Both bind cdc2 kinase, but A-type complexes are activated much earlier in the cell cycle than B-type complexes, and only B-type cyclins are found in purified MPF.

To investigate the role of cyclin A, we utilized extracts from metaphase-arrested eggs that are able to exit mitosis *in vitro*, to undergo DNA synthesis, and then to reenter mitosis. These extracts retain the characteristic dependence of mitosis upon completion of DNA synthesis—that is, will not enter mitosis if DNA synthesis has not been completed, which can result from an excess of DNA in the system or the presence of aphidicolin, an inhibitor of DNA polymerase. By using antisense oligodeoxynucleotides to ablate cyclin A mRNA from the system, we were able to show that activation of MPF (cyclin B/cdc2) occurred even when DNA synthesis had not been completed. Readdition of recombinant cyclin A protein to the antisense-ablated extracts restored the dependence of mitosis on DNA synthesis by causing a lengthening of S phase until DNA synthesis was complete.

This provides evidence that one of the functions of cyclin A is in the crucial checkpoint that prevents the activation of cyclin B/cdc2 (MPF) until DNA synthesis has been completed. It also explains why cyclin A/cdc2 complexes are activated earlier in the cell cycle than cyclin B/cdc2 complexes and why cyclin A is degraded before cyclin B. The latter phenomenon, long known, is clearly appropriate, since cyclin A exerts an inhibitory function on cyclin B. A question that now merits attention concerns what substrates exist for phosphorylation by cyclin A/cdc2 that are involved in this feedback control mechanism.

We are interested in the mechanism of activation of MPF in oocytes during the cell cycles of meiosis I and II. In these cycles the synthesis of proteins other than cyclin are required for MPF activation. One protein required for meiosis I and II is the product of the mos proto-oncogene. Proto-oncogenes are the normal cellular counterpart of mutated oncogenes found in cancer cells,

suggesting that oncogenes act by perturbing normal cellular pathways. In general, very little is known about how proto-oncogenes work, but the specific involvement of mos in cell cycle control is the clearest example of a specific function for any proto-oncogene in a defined cellular process. The *mos* gene encodes a serine/threonine protein kinase, indicating the existence of a substrate for phosphorylation by mos that can lead to activation of MPF as well as stabilization of cyclin in the metaphase arrest of meiosis II. The transition between meiosis I and II is also perhaps the only well-documented case in which cdc2 kinase activity declines without accompanying cyclin B degradation.

This year Linda Roy observed that introduction of protein synthesis inhibitors into the system at the transition between meiosis I and II leads to the immediate destruction of cyclin B, which would otherwise be stable, thereby establishing that the stability of cyclins between meiosis I and II requires continuous synthesis of protein. Initially we expected that the protein synthesis required for cyclin stability would be exemplified by the synthesis of the mos proto-oncogene kinase itself, since that kinase is known to comprise one of the major newly synthesized proteins during maturation. However, Dr. Roy found that, in fact, the mos kinase remains fully active even when B-type cyclins are degraded. This implies that there is another protein whose synthesis is required for cyclin stability besides the mos proto-oncogene.

Dr. Roy also did the converse experiment by microinjecting antisense oligodeoxynucleotides against the mos kinase mRNA into cells just prior to entry into meiosis I. By this procedure she was able to ablate greatly the level of mos proto-oncogene expression. However, under these conditions, cyclin B does not undergo degradation between meiosis I and II, thereby establishing that mos is not necessary for the stability of cyclin B between meiosis I and II. These results suggest that other components besides the mos kinase are required for the unusual cyclin stability between meiosis I and II, and clearly the identification of these other components merits attention.

$G_1 \rightarrow S$ Regulation

The cell cycle restriction point in G_1 governing the $G_1 \rightarrow S$ transition that involves cdc2-like kinases has been termed START in yeast and the R point in mammalian cells. In both budding yeast and fission yeast, it is quite clear that the genuine *cdc2* gene product mediates both the G_1 and G_2 control points. Recently it has become evident that the regulation of the $G_1 \rightarrow S$ transition in vertebrate cells is considerably more complicated than in yeast. One aspect of this complexity involves the presence of another *cdc2*-like gene in the G_1 phase that may mediate events at the $G_1 \rightarrow S$ transition. This gene was originally discovered in *Xenopus* and given the name Eg1 but has since been renamed cyclin-dependent kinase 2, or cdk2. The idea that cdk2 might be a form of cdc2 specialized for G_1 control has come from the finding by others that cdk2 itself will not complement mutations in the *Saccharomyces cerevisiae cdc2* cognate gene *CDC28* that affect the $G_2 \rightarrow M$ restriction point, but will substitute at least partially for the G_1 function of *CDC28* when coexpressed with a G_1 cyclin from human cells.

This year we have made a major effort to study the biochemistry and regulation of cdk2 in the *Xenopus* embryonic cell system. Toward this end we have developed an antibody against the protein and used it to show that the protein kinase activity of cdk2 oscillates in the cell cycle with a periodicity similar to cdc2. In the *Xenopus* embryonic cell cycle, cdk2 is not associated with either cyclin A or cyclin B, but instead with two proteins of 36 and 48 kDa. This suggested the possibility that cdk2 would be regulated by phosphorylation, perhaps in ways similar to the known regulation by phosphorylation of cdc2 kinase itself.

Recently we showed that in fact cdk2 is a phosphoprotein, that it is phosphorylated on tyrosine and serine residues, and that phosphorylation of these residues changes during the cell cycle. In particular, the presence of phosphotyrosine correlates with less-active forms of the protein, but the major site of phosphorylation was determined by tryptic phosphopeptide mapping to be in a phosphopeptide distinct from the one containing tyrosine 15, so characteristic of cdc2 kinase. Despite this difference in the sites of phosphorylation, however, it would appear that the same enzyme responsible for the dephosphorylation of cdc2 is also active on cdk2.

By continuing to study both cdc2 kinase in the $G_2 \rightarrow M$ transition and cdk2 kinase in the $G_1 \rightarrow S$ transition, it should be possible to achieve a comprehensive understanding of cell cycle control at the two main restriction points present in eukaryotic cells.

The Role of T Cells in Health and Sickness

Philippa Marrack, Ph.D.—Investigator

Dr. Marrack is also a member of the Division of Basic Immunology of the Department of Medicine at the National Jewish Center for Immunology and Respiratory Medicine, Denver, and Professor of Biochemistry, Biophysics and Genetics, of Microbiology and Immunology, and of Medicine at the University of Colorado Health Sciences Center, Denver. She took her Ph.D. in biological sciences at the MRC Laboratory for Molecular Biology in Cambridge, England, and then did postdoctoral work with Richard Dutton at the University of California, San Diego. From there she moved to the University of Rochester, and, after seven years, to her present position. Dr. Marrack is a member of the National Academy of Sciences and was recently awarded the Christopher Columbus Discovery Award for Biomedical Science.

T cells are essential to the ability of higher vertebrates to resist disease. These cells are not only able to destroy invading organisms by killing cells in which such organisms live, but they also produce hormone-like substances that contribute to protection against invasion, by stimulating production of protective antibodies, for example.

T cells bear receptors for antigen on their surface. There are about 20,000 receptors on each cell. These are composed of two polypeptide chains, α and β, each made up of a number of variable elements: Vα, Jα, Vβ, Dβ, and Jβ. In each T cell the receptors are composed of different combinations of these elements, so that each T cell has receptors with a somewhat different structure.

As far as we know, receptors are assembled randomly from the available Vαs, Jαs, and so on, while T cells are developing. Consequently there is a distinct possibility that some T cells will bear receptors able to interact with self antigens, i.e., components of the individual containing the T cells. These cells are a potential threat, since they could attack and destroy their own host. Usually, however, they themselves are destroyed or inactivated before they become mature enough to cause damage.

There is reason to believe that some potentially self-reactive T cells avoid the processes of tolerance and are allowed to mature. Even though they could attack tissues of their host, they seem not to, perhaps because the self antigens with which they could interact are sequestered in a tissue that they cannot reach. Occasionally, however, these self-reactive cells are activated by encounter with an environmental antigen. After stimulation the cells become more motile and active and are then able to interact with and destroy the tissues of their host, causing a so-called autoimmune disease.

A number of human diseases—juvenile diabetes, lupus erythematosus, and multiple sclerosis among others—may be due to T cell malfunction similar to that described above. Our own work, in collaboration with Brian Kotzin (Department of Pediatrics, National Jewish Center), has concentrated on rheumatoid arthritis. About 2 million people in the United States suffer from this disease, which seems to be due to an immune attack on material in joints. We have shown that many of those afflicted have unexpectedly low levels of T cells bearing Vβ14 in their blood, and some of the missing cells are present in the fluid bathing the rheumatic joints.

We have suggested that rheumatoid arthritis involves chronic invasion of the host by a foreign antigen able to interact with T cells bearing Vβ14. This antigen would first stimulate target T cells, and then cause their disappearance, probably by the processes of tolerance described above. Perhaps the subset of Vβ14-bearing cells that can recognize self antigens is waylaid and rescued from death by sequestration in the joints. Here they cause inflammation and the symptoms of arthritis.

At the moment all these are simply ideas upon which to base future experiments on the cause, treatment, and prevention of rheumatoid arthritis. Meanwhile, we have begun a search for the foreign antigen that may interact with Vβ14-bearing T cells and thus start the disease. There is every indication that this is a superantigen, a special kind of antigen that interacts with T cells, primarily via the Vβ portion of their receptors. It is also likely, as mentioned above, that the antigen is produced by a chronic infectious agent, such as a chronic virus.

Until recently the only infectious agents known to produce superantigens were mycoplasma and bacteria, such as staphylococci and streptococci. These agents do not usually infect their hosts in a chronic fashion. Early in 1991, however, we and several other groups showed that certain viruses, a collection of retroviruses present in mice, can also express superantigens. Encouraged by this result, we decided to screen human viruses for such expression. Our preliminary results suggest that Epstein-Barr virus, the

cause of a common chronic infection and of infectious mononucleosis in humans, may indeed encode a superantigen. While we do not think that Epstein-Barr virus is the causative agent of rheumatoid arthritis, it is possible that other members of the Herpesviridae family may be.

Cell Regulation by Transforming Growth Factors

Joan Massagué, Ph.D.—Investigator

Dr. Massagué is also a member of the Cell Biology and Genetics Program at Memorial Sloan-Kettering Cancer Center and Professor of Cell Biology at Cornell University Graduate School of Medical Sciences, New York. He received his Ph.D. degree in biochemistry from the University of Barcelona, Spain, and was a postdoctoral fellow with Michael Czech at Brown University. He was Assistant and Associate Professor of Biochemistry at the University of Massachusetts Medical School before assuming his present position.

THE proliferation of cells is controlled by a balance of positive and negative signals. The machinery that conveys growth inhibitory signals is similar in design to that which signals cell growth. Both involve 1) factors that circulate between cells and 2) membrane receptors that are coupled to signal transduction circuitry inside the cell. The signals carried by growth-promoting factors have been extensively studied for the past two decades. The growth inhibitors, however, have come to the attention of biologists only recently. Yet they include some of the most widespread and versatile regulators of cell growth and phenotype. Some of them are implicated in processes of development, tissue repair, and recycling, and their study may show us ways to constrain the unrestricted growth of cancer cells.

Multifunctional Growth Inhibitors and Their Receptors

Ranking high in the list of growth inhibitors, and in our research interests, is the polypeptide TGF-β (transforming growth factor-β). In reality, TGF-β represents a large family of growth and differentiation factors that also includes the activins, the bone morphogenetic proteins, the Müllerian inhibiting substance, and others. The evolutionary conservation of these factors is unusually strict, and they are broadly multifunctional. For example, TGF-β can inhibit cell proliferation, regulate cell differentiation, affect how cells organize tissue structures, and perform various other functions in cells from virtually every lineage. Likewise, other members of the family, such as activin and decapentaplegic, are involved in body axis formation during embryo development.

Over the past year our research program has centered on identifying and isolating genes that encode receptors for TGF-β and related factors. We had previously determined that TGF-β binds to various types of receptor proteins on the cell membrane, and we have recently cloned genes for two classes of such receptors. One class comprises receptor membrane proteins that can trigger a biochemical reaction inside the cell upon binding a factor on the outside. The receptors in this class probably mediate most cellular responses to these factors. Like the factors they bind, these receptors exist in many variants, each probably representing a discrete adaptation to achieve optimal control of cell functions.

One member of this receptor class binds the TGF-β–related factor activin. The biochemical reaction carried out by this receptor is the transfer of phosphate from ATP to serine and threonine groups present in certain intracellular proteins. Phosphorylation of serines and threonines represents a significant departure from previously known receptor signals. In contrast to the receptors for activin and TGF-β, those for growth-promoting factors typically trigger phosphorylation of tyrosine groups in proteins. Thus the tyrosine phosphorylation signals of mitogens are challenged by the serine/threonine phosphorylation signals of antimitogens. These findings offer new insight into the counteracting signals that preserve the normal balance of cell growth. This work has been supported by a grant from the National Cancer Institute.

Receptor Accessory Molecules

The other TGF-β receptor class recently cloned by our group is interesting for other reasons. This protein, called betaglycan, is thought to act as a helper of the signaling receptors. Rather than mediate cell responses directly, betaglycan seems to regulate the access of cells to TGF-β by either helping present this factor to the signaling receptors or storing it for later use by the cell.

The structure of betaglycan is unusual for a growth factor–binding protein. It consists of a core protein that carries a large mass of negatively charged carbohydrate. TGF-β binds the core protein, whereas the carbohydrate can bind the so-called heparin-binding growth factors. Work is under way to map the portion of this molecule that binds TGF-β and to test its ability as a modulator of TGF-β activity.

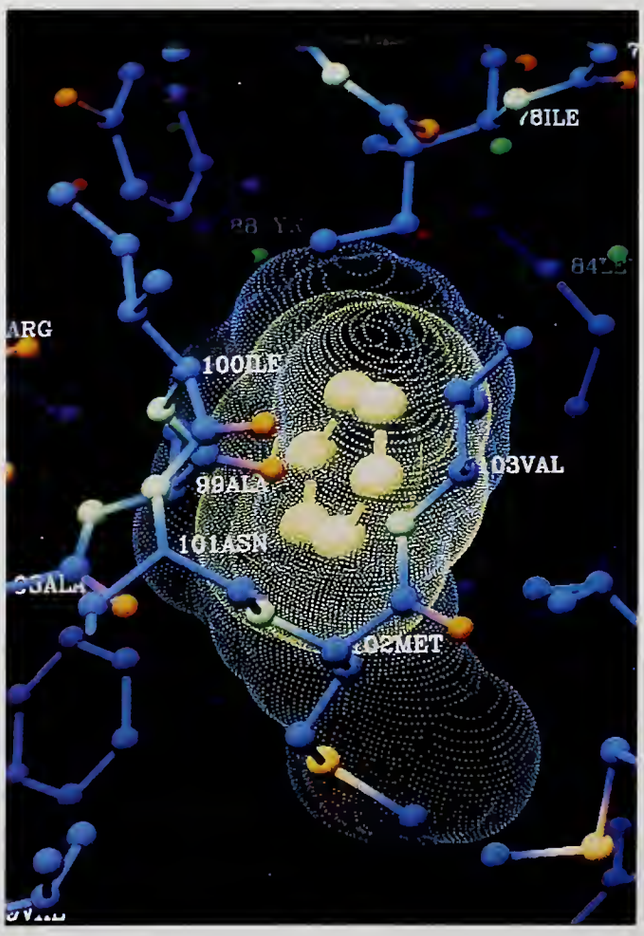
With the cloning of these genes, it is now possi-

ble to explore their properties in detail and to identify key pieces of the machinery that transfers growth inhibitory signals from the membrane to the nucleus. Furthermore, it should be possible to determine to what extent the genetic loss of TGF-β receptors might cause loss of constraint in cell proliferation and thus incite tumor cell outgrowth.

Cell-Cell Stimulation by Membrane-bound Growth Factors

Paracrine growth factors and polypeptide hormones are generally synthesized as larger soluble precursors that are later fragmented to yield the bioactive forms. In a recently found variation of this theme, factors such as TGF-α (no structural relationship to TGF-β) are generated from membrane-anchored proteins rather than from soluble precursors. The TGF-α precursor can accumulate on the cell membrane and bind to receptors located on the surface of adjacent cells. This interaction can sustain cell-cell adhesion and stimulate DNA replication by cell-cell contact.

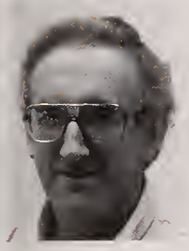
In work supported by a grant from the National Cancer Institute, we have shown that generation of TGF-α by cleavage of its precursor occurs at the cell surface by a highly regulated enzymatic system. This system is strongly activated by tumor-promoting phorbol esters and growth factors via mechanisms involving protein kinase C and calcium influx into the cell. Thus the precursor cleavage process functions as a regulated switch between two active forms of the growth factor, one membrane-bound and the other diffusible. The membrane-bound forms could be important in tissue development processes whose guidance depends on discrete cell-cell interactions incompatible with the diffusible nature of soluble factors. The regulated nature of the pro-TGF-α cleavage process renders it susceptible to exogenous control with pharmacologic agents. Furthermore, it provides a way to identify the pieces of the general machinery controlling the release of this and other membrane proteins that mediate cell-to-cell interaction.



In this molecular model, a benzene molecule (in yellow) is shown buried within the core of T4 lysozyme in a cavity created by replacing leucine 99 with alanine. The inner dotted form represents the van der Waals surface of the benzene, and the outer envelope, the van der Waals surface of the cavity.

Research and photograph by A. Elisabeth Eriksson and Xue-jun Zhang in the laboratory of Brian Matthews.

Structural Basis of Interactions Within and Between Macromolecules



Brian W. Matthews, Ph.D., D.Sc.—Investigator

Dr. Matthews is also Professor of Physics and Director of the Institute of Molecular Biology at the University of Oregon and Adjunct Professor of Biochemistry at the Oregon Health Sciences University, Portland. He received his undergraduate and graduate training at the University of Adelaide, Australia. He did postdoctoral research at the MRC Laboratory of Molecular Biology, Cambridge, England (with David Blow) and at the National Institutes of Health (with David Davies). Dr. Matthews is a member of the National Academy of Sciences.

OUR laboratory uses x-ray crystallography, in concert with other techniques, to try to address some of the fundamental problems in biology: How do proteins spontaneously fold into their biologically active three-dimensional configurations? What determines the stability of these folded proteins, and can stability be improved? How do proteins interact with each other? How do they interact with DNA? How do enzymes act as catalysts?

The Protein-folding Problem

An area of long-standing interest is the so-called protein-folding problem. How does a newly synthesized, extended peptide chain "know" how to fold spontaneously into its active three-dimensional shape?

Although it has long been recognized that the amino acid sequence of a protein determines its three-dimensional structure, recent work from several laboratories has made it clear that certain amino acids are more important than others in the folding process. At some positions, typically the solvent-exposed mobile sites in the folded protein, amino acids can be interchanged almost at random with little apparent effect on folding or stability. On the other hand, interchange of amino acids in buried or rigid parts of a folded protein can destabilize it, suggesting that the amino acids at these positions are important in determining the folded conformation.

One of the encouraging developments has been the relative freedom with which amino acid replacements can be introduced in a protein of interest. To try to simplify the complexity of the protein-folding problem, we are attempting to replace some of the "nonessential" amino acids in phage T4 lysozyme with alanine. Such a "polyalanine protein" would, in principle, truncate all nonessential side chains and allow one to focus on those parts of the amino acid sequence that are critical for the folding process.

In experiments to date, a series of alanines has been introduced within two different α-helices of T4 lysozyme. The somewhat surprising result is that alanines are not only tolerated at most positions in the α-helix; they can sometimes increase the protein's stability. In an extreme case it has been found that 10 alanines can be introduced in sequence, yet the protein still folds normally and has full activity. This illustrates that the information in the amino acid sequence of a protein is highly redundant.

Understanding the Interactions That Stabilize Protein Structures

It is generally agreed that the major factor in stabilizing the folded structures of globular proteins is the hydrophobic effect. Until recently it has also been generally agreed that the strength of the hydrophobic effect—i.e., the energy of stabilization provided by the transfer of hydrocarbon surfaces from solvent to the interior of a protein —is about 25–30 cal mol^{-1} for each square angstrom of surface area buried within the protein. However, some recent studies using site-directed mutagenesis and protein denaturation have suggested that the strength of the hydrophobic effect might be much higher.

A principal difficulty in addressing this problem has been the lack of relevant structural data. How does a protein structure respond when a bulky hydrophobic residue such as leucine is replaced by a smaller residue such as alanine? Does the protein structure remain essentially unchanged or is there structural rearrangement to avoid the creation of a cavity? If cavities are created, do they contain solvent?

To address these questions, six "cavity-creating" mutants in which a large hydrophobic amino acid was replaced by a smaller one were constructed within the hydrophobic core of phage T4 lysozyme. All variants were crystallized and the structures determined at high resolution.

The structural consequences of the mutations differ from site to site. In some cases the protein structure hardly changes at all. In other cases, however, both side-chain and backbone shifts up to 0.8–1.0 Å were observed. In every case removal of the wild-type side chain allowed some of the surrounding atoms to move toward the vacated space, but a cavity always remained.

This suggests a way to reconcile the different values for the apparent strength of the hydrophobic effect. One can imagine two extreme situations. In one case a leucine → alanine replacement is constructed, and the protein structure remains completely unchanged. In this situation the size of the created cavity is large, and the mutant protein is maximally destabilized. In the other extreme, the protein structure relaxes in response to the leucine → alanine substitution, fills the space occupied by the leucine side chain, and so avoids the formation of any cavity whatsoever. In this case the decrease in energy of the mutant protein relative to wild type drops to a constant energy term that is characteristic for a leucine → alanine replacement.

Ligand Binding Within Cavities

We have shown by crystallographic and thermodynamic analysis that the cavity created by the replacement leucine 99 → alanine in T4 lysozyme is large enough to bind benzene and that ligand binding increases the melting temperature of the protein by 5.7°C. This shows that cavities can be engineered within proteins and suggests that such cavities might be tailored to bind specific ligands. The binding of benzene at an internal site 7 Å from the molecular surface also illustrates the dynamic nature of proteins, even in crystals.

Receptor-Ligand Interaction

To develop an understanding of the mode of action of growth factors and their interactions with their receptors, we have crystallized and determined the high-resolution structure of human fibroblast growth factor. The structure is very similar to that of interleukin-1β. Clearly, many growth factors have similar overall structures, but the exact relationship of these factors in the vicinity of their receptor-binding regions remains to be clarified. Structural studies of human nerve growth factor are in progress.

Protein-DNA Interaction

We have been interested for some time in the interaction between proteins and nucleic acids. In 1981 we determined the structure of the Cro repressor protein of λ bacteriophage (bacteria-infecting virus), one of the prototypical examples of a DNA-interacting protein. The structure of Cro, as determined crystallographically, suggested that a characteristic part of the protein, now known as the helix-turn-helix motif, is especially important in DNA binding. The helix-turn-helix unit can be considered as a "reading head" that fits into the grooves of the DNA and matches the DNA structure at the specific recognition site. The helix-turn-helix motif is now known to occur in a large number of DNA-binding proteins, and its functional role has been confirmed by structures of several DNA-protein complexes.

We have subsequently determined the crystal structure of Cro protein in complex with a tight-binding 17–base pair DNA operator. In general terms the structure of the complex supports the model for Cro-DNA interaction that was proposed on the basis of the uncomplexed protein, although the Cro dimer undergoes a substantial conformational change relative to the uncomplexed crystal structure.

Recently we have determined the structure of the biotin repressor from *Escherichia coli*. This is a more complicated protein that requires the presence of an effector molecule to bind DNA. It also acts as an enzyme.

Studies of protein stability and protein-DNA interaction were supported in part by grants from the National Institutes of Health.

What Viruses Are Telling Us About Gene Regulation in Mammalian Cells

Steven Lanier McKnight, Ph.D.—Investigator

Dr. McKnight is also a staff member in the Department of Embryology at the Carnegie Institution of Washington, Baltimore, and Adjunct Professor in the Departments of Biology and of Molecular Biology and Genetics at the Johns Hopkins University School of Medicine. He earned his Ph.D. degree in biology from the University of Virginia and, except for four years with the Fred Hutchinson Cancer Research Center in Seattle, has been with the Carnegie Institution ever since. Dr. McKnight was recently elected to the National Academy of Sciences.

VIRUSES that attack mammalian cells rely on preexisting enzymes, factors, and cellular functions to negotiate their infectious cycle. Molecular studies of virus infection have thereby provided key insights into normal cellular processes. For example, studies of the processing and transport of membrane glycoproteins that form the exterior coats of influenza virus and vesicular stomatitis virus have helped to explain how proteins are selectively transported to the appropriate cellular compartment.

Studies of viruses have also illuminated complex phenomena regarding selective gene expression. RNA splicing was first discovered in studies of human cells infected by adenovirus. Likewise, the capacity of DNA segments known as enhancers to regulate gene expression from remote locations was discovered in studies of simian virus 40.

Work from our laboratory has focused on the mechanisms of herpesvirus gene regulation. The herpesvirus chromosome contains roughly 50–100 genes that are expressed in a tightly controlled temporal cascade. Early during the infectious cycle, five immediate-early (IE) genes are activated. The IE genes encode protein products, termed transcription factors, that act to regulate subsequent viral gene expression. Several hours later about 25 delayed-early (DE) genes are activated. Transcription of DE genes is strictly dependent on the prior production of IE proteins. DE genes encode proteins required for replication of viral DNA. Following viral DNA replication, about 25 late (L) genes are activated. Herpesvirus L genes encode structural proteins that form the intact virus particle, including a complex set of membrane glycoproteins and structural proteins of which the viral capsid is composed.

Interestingly, one of the L gene products encapsidated in the mature virus is a potent and specific transcription factor dedicated to the activation of IE genes in the subsequent infectious cycle. Roughly 1,000 molecules of this protein, termed viral protein 16 (VP16), are packaged into the mature virus particle. Upon infection of an otherwise healthy cell, VP16 is released from the infecting virus and comes to be associated with enhancer elements located upstream from each IE gene. Thus emplaced, VP16 acts as a potent "trigger" for the rapid and prolific expression of IE genes.

Early studies in our laboratory were focused on the herpesvirus DE gene encoding thymidine kinase (TK). In order to probe the mechanisms of activation by IE proteins, the TK gene was systematically mutated with the aim of defining specific regulatory switch points within or around the gene. A region of 35 base pairs encompassing the site of transcription initiation was identified as being responsive to activation by the IE transcription factors. Surprisingly, however, such studies identified three additional regulatory DNA sequences located upstream of the TK gene. These supplementary regulatory elements were identified as binding sites for host cell proteins, including the Sp1 (selectivity protein 1) transcription factor discovered by Robert Tjian (HHMI, University of California, Berkeley) and his colleagues.

The involvement of host cell transcription factors in viral gene expression has since been observed in numerous cases. An exciting recent discovery along such lines has come from Joseph Nevins (HHMI, Duke University Medical Center). Dr. Nevins and his colleagues have identified a human transcription factor, termed E2F (early region 2 factor), that plays a pivotal role in the transcriptional induction of certain adenovirus genes. Remarkably, the E2F factor has been shown to form a specific complex with the product of a recessive oncogene encoded by the retinoblastoma locus. Such observations are beginning to provide mechanistic insight into the role of the retinoblastoma protein in growth control and cancer.

More recent studies in our laboratory have focused on the activation of herpesvirus IE genes during the early stage of infection of cultured mammalian cells. As mentioned previously, IE gene activation is stimulated by VP16, a virally encoded L protein that is encapsidated in mature virus particles. Like the soldiers sequestered in

the proverbial Trojan horse, VP16 molecules are deposited in the newly infected cell as "ready-made" transcription factors. In an elegant series of experiments, Bernard Roizman and his colleagues at the University of Chicago showed that the activating specificity of VP16 for herpesvirus IE genes is dictated by regulatory DNA sequences located upstream of each IE gene. Such IE "enhancers," when linked onto another gene otherwise unresponsive to VP16, conferred a direct and specific response.

On its own, VP16 is incapable of direct association with the IE enhancers. In order to develop an understanding of the molecular mechanisms controlling transcriptional activation by VP16, efforts have been undertaken to identify and purify cellular proteins that bind to IE enhancers. At least four different host cell proteins are required for maximal activation of IE gene expression by VP16. The most recent work in our laboratory has entailed the identification and purification of a DNA-binding complex that associates avidly and specifically with a critical switch point within IE enhancers. This particular switch point, termed in the jargon of the transcription field a "cis-regulatory element," is largely composed of guanine (G) and adenine (A) residues. The cellular DNA-binding protein that interacts specifically with the GA-rich cis-regulatory element has been termed GA-binding protein (GABP).

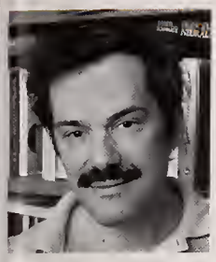
Kelly LaMarco, a postdoctoral fellow in our laboratory, purified GABP and found it to be composed of two polypeptide subunits. After deriving the partial amino acid sequence from proteolyzed fragments of each subunit, Dr. LaMarco, along with postdoctoral associates Catherine Thompson and Tom Brown, succeeded in cloning the genes encoding each GAPB subunit. The amino acid sequence of one subunit, termed GABPα, exhibits significant similarity to the product of the ETS proto-oncogene. Indeed, the ETS-related region of GABPα represents the part of the protein that mediates direct contact with DNA.

This same region of the protein is also necessary for protein-protein contact with the other subunit of the complex, termed GABPβ. Fortunately, the sequence of GABPβ also exhibits a region of amino acid sequence similarity with previously studied proteins. GABPβ contains four imperfect repeats, 33 amino acids in length, that are related to similarly sized repeats present in the products of a number of interesting proteins.

One such relative is the product of the *Notch* gene of fruit flies studied by Spyridon Artavanis-Tsakonas (HHMI, Yale University). The *Notch* gene product is a membrane protein that plays an important role in cell-cell communication during fruit fly development. Another protein that contains the 33 amino acid repeats is ankyrin, a cytoskeletal protein in red blood cells discovered by G. Vann Bennett (HHMI, Duke University Medical Center). Drs. Thompson and Brown found that the 33 amino acid repeats of GABPβ constitute the part of the protein required for direct interaction with its matching subunit (GABPα), thus providing the first conclusive evidence for the mechanist role of this protein structural motif.

Many questions regarding the properties and function of this protein complex remain unresolved. For example, how does binding of the complex facilitate activation of herpesvirus IE genes? What role does GABP play in the control of cellular gene expression? Might GABP in some way influence the decision of herpesvirus to execute its lytic cycle (as it does in epithelial cells) or the latent state it enters when the virus infects neuronal cells? Given a critical set of molecular reagents, including recombinant DNA clones and specific antibodies, it should now be possible to address these questions directly.

Fundamental Mechanisms of Ion Channel Proteins



Christopher Miller, Ph.D.—Investigator

Dr. Miller is also Professor of Biochemistry at Brandeis University and Adjunct Professor of Molecular Biology at Massachusetts General Hospital, Boston. He received his B.A. degree in physics from Swarthmore College and his Ph.D. degree in molecular biology from the University of Pennsylvania. He carried out postdoctoral work in membrane biochemistry with Efraim Racker at Cornell University for two years and then joined the Graduate Department of Biochemistry at Brandeis.

ION channels are the most fundamental elements of molecular hardware in the nervous system. They are the membrane-spanning proteins that directly mediate the transmembrane ionic fluxes by which electrical signals are generated, propagated, and integrated in neurons, muscle, and other electrically active cells. By forming aqueous pores through the heart of the channel protein (and hence through the membrane that the protein spans), channels act as "leakage" pathways for ions down their preestablished thermodynamic gradients. These proteins make intelligent leaks. Channels can discriminate fiercely among the different species of inorganic ions present in the aqueous solutions bathing the cell membrane. They can also rapidly open and close their conduction pores in response to external signals, such as binding of neurotransmitters or changing of the transmembrane electric field.

Work here is directed toward questions of basic molecular mechanisms of ion channel operation and of the underlying protein structures. Since no high-resolution structures have been obtained (or are soon likely to be) for this class of proteins, it is necessary to draw structural inferences from close examination of ion channel function. This can be done because ion channels, unique among proteins, can be studied quantitatively at the single-molecule level. In this laboratory, heavy use is made of the technique of "single-channel reconstitution," in which individual ion channel molecules are inserted into an artificial membrane under simple, chemically controllable conditions.

This approach has allowed us to develop crude physical pictures of several ion channels in which crucial dimensions have been deduced: the conduction pore's width and length, the distance of its entryway from the lipid bilayer surface, and the number of ions inside the channel during the conduction process. We are currently complementing these purely functional and mechanistic studies with recently developed methods of membrane protein biochemistry and manipulation of ion channels at the genetic level.

Use of Peptide Neurotoxins as Probes of K^+ Channel Structure

Charybdotoxin (CTX) is a scorpion venom-derived peptide that blocks a small family of K^+-specific channels. Having shown that it acts by physically plugging the channel's externally facing mouth, we are now utilizing CTX as a probe of this important region. We are using two different K^+ channels for these efforts: the high-conductance Ca^{2+}-activated K^+ channel from skeletal muscle reconstituted into planar lipid bilayers, and the Shaker K^+ channel expressed in *Xenopus* oocytes. Each of these channels has a particular advantage, the former lending itself to close mechanistic study and the latter to molecular genetic manipulation.

Using a synthetic gene for CTX expressed in *Escherichia coli*, in combination with the known structure of the peptide, we have mapped the interaction surface of the CTX, using both types of K^+ channels as CTX receptors. As a result, we know that the peptide makes intimate contact with the channel on a well-defined surface built from 7–8 residues. One of these residues interacts electrostatically with a K^+ ion residing in the channel's conduction pore.

In parallel with mapping the toxin, we have used site-directed mutagenesis with the Shaker K^+ channel to identify channel residues making up the CTX receptor located in the channel's external mouth. With this information in hand, we are currently trying to identify channel-toxin interaction partners by constructing complementary mutants of peptide and receptor. This will allow us to deduce physical distances between residues lining the mouth of the K^+ channel.

Purification and Reconstitution of Cl^- Channels

The electric ray *Torpedo californica* carries in its electric organ a Cl^--specific channel with an unusual structural characteristic. The channel is built as a dimeric, or "double-barreled" complex, with two identical Cl^- diffusion pathways in a single molecular unit. We have developed a functional assay for this channel protein in a solu-

bilized state and are performing conventional membrane purification studies. The channel is expressed in its natural tissue at high density, so milligram quantities should be easily obtained. The isolation of this channel will enable us to study the molecular family of voltage-dependent anion channels at the protein-biochemical level.

Structure-Function Relations in a Minimal K^+ Channel

We are beginning a structure-function analysis on a K^+-specific channel, first cloned from a kidney cDNA library, that shows a remarkable molecular property: a very small polypeptide size of only 130 amino acids, some 10-fold smaller than, for example, the voltage-dependent Na^+ channel. We have constructed and expressed a fully synthetic gene for this channel, using the degeneracy of the genetic code to build a large number of unique restriction sites throughout the coding sequence. Thus prepared to perform routine cassette mutagenesis, we are initiating a search for functional domains of the channel and developing direct tests to determine whether this gene does in fact code for an ion channel at all, a basic question that has yet to be answered rigorously.

We have found point mutations in the single hydrophobic domain that specifically alter ion selectivity among close K^+ analogues as well as the affinities of several channel blockers. The results not only demonstrate that this cDNA does code for a structural gene for a K^+ channel but also provide initial intimations about the way the K^+-conducting pore is formed.

Neural Foundations of Vision

J. Anthony Movshon, Ph.D.—Investigator

Dr. Movshon is also Professor of Neural Science and Psychology at New York University and Adjunct Professor of Physiology and Biophysics at New York University Medical Center. He received his B.A. degree and his Ph.D. degree in experimental psychology from Cambridge University, where he worked with Colin Blakemore. After joining the faculty of NYU, Dr. Movshon has remained there except for a sabbatical year at Oxford University. He was founding director of the NYU Center for Neural Science. Among his honors are the Young Investigator Award from the Society for Neuroscience and the Rank Prize in Optoelectronics.

OUR research concerns the function and development of the visual system, especially the visual areas of the primate cerebral cortex. Our main experimental tools are electrophysiological recording and quantitative analysis of the visually evoked activity of single neurons. We also draw importantly from related work in visual psychophysics, computational modeling, and complementary neuroanatomy.

Presently we are involved in two broad groups of studies. The first concerns the functional properties of single neurons in the extrastriate visual areas of the macaque monkey's cerebral cortex, with special emphasis on the processing of information about visual motion, space, form, and color. The second group of studies concerns the development of cortical visual function in monkeys and the way that development is affected by abnormal early visual experience.

An important organizing theme derives from the discovery of two functional streams in the monkey's geniculo-cortical visual pathway. One stream, the P system, originates in the dense and numerous Pβ ganglion cells of the retina, continues through the parvocellular layers of the lateral geniculate nucleus (LGN), and extends to layer 4Cβ of the striate (or primary visual) cortex, V1. A second stream, the M system, originates in the large, fast-conducting but relatively sparse Pα retinal ganglion cells, continues through the magnocellular layers of the LGN, and enters the striate cortex through layer 4Cα.

Signals from the P system are passed preferentially into a set of cortical areas that seem to be of special importance for the processing of form and color, especially visual areas 2 and 4, and into the inferior temporal cortex. Signals from the M system pass rather selectively into another set of cortical areas that seem to be essential for the analysis of visual motion and visual space, especially the middle temporal area (MT or V5), and into the posterior parietal cortex. Our working hypothesis is that these streams subserve different, albeit overlapping visual functions, and also that different forms of developmental visual disorder may reflect abnormalities primarily affecting one stream or the other.

To study the functions of cortical visual areas, we analyze the responses evoked in single neurons by visual stimuli carefully selected to permit formal characterization of underlying neuronal mechanisms. The class of properties in which we are generally interested concerns the selectivity with which neurons respond to variations along one or another visual dimension. We also try to examine the neuroanatomical distribution and functional properties of neurons providing afferent signals to a particular area, so that we can attempt to understand the computational transformations of the visual signal executed by the circuits in each area.

An important concern is to establish the particular dimensions of the visual stimulus for which neurons in that area show an invariant selectivity —that is, for which their selectivity is unaffected by parametric variation in other, unrelated dimensions. For example, neurons in V1 have invariant selectivity for the spatial, temporal, and chromatic structure of visual stimuli. Neurons in MT transform afferent spatiotemporal signals into invariant representations of an object's speed and direction. Neurons in V4, on the other hand, may transform simple afferent chromatic signals into invariant representations of the object's surface properties.

A critical issue in cortical sensory physiology is to relate perceptual experience and judgment to the activity of neurons and neuron assemblies. In collaboration with William Newsome at Stanford University, we have used statistical methods based on the theory of signal detection to compare the performance of single neurons with psychophysical measures of performance obtained concurrently from an awake, behaving monkey. The goal is to deduce the associations between the computation of perceptual features and the activity of particular groups of neurons. The results suggest that small groups of neurons in area MT may carry the signals upon which behaving monkeys make judgments of the motion content of visual targets. This approach allows us to form a common language in which to consider psychophysical, computational, and neurobiological

analyses of visual cortical function. We are also applying methods derived from statistical decision theory to the question of the absolute efficiency of visual representations in the brain.

In addition to its purely visual functions, the M stream provides signals that drive eye movements of pursuit, the slow, smooth eye movements with which primates stabilize on the retina the image of a moving visual target. In collaboration with Stephen Lisberger (University of California, San Francisco), we have studied several aspects of the relationship between visual and visuomotor processes. In a series of neurophysiological studies, we have explored the responses of MT neurons to the dynamic motion profiles used to characterize pursuit and have documented the suitability of the motion-related signals in these neurons for the task of initiating pursuit. In psychophysical work, we have begun to explore the kinds of visual signals that pass into the oculomotor system by examining the relationship between the detectability of particular visual patterns and the pursuit eye movements they elicit.

Our overall ambition for these studies is to "turn the sensory-motor corner" and relate the particulars of visual processing to the higher mechanisms that produce voluntary motor commands. To this end, we are developing computational models designed to explain the signal transformations that take place at a series of stages between the initial registration of the visual image and the formulation of the final oculomotor command.

To analyze development, in a project supported by the National Eye Institute, we study the vision of monkeys reared either with an artificial strabismus (deviation of one eye) or anisometropia (difference in the refractive state of the two eyes). Both of these manipulations lead to conditions resembling human amblyopia, a common visual deficit of central nervous system origin. In behavioral experiments, we learn how experimental amblyopia affects perceptually defined mechanisms that support visual sensitivity to form, contrast, and position. Neurophysiological studies in the same animals then reveal alterations in cortical neuron properties that seem to be related to the psychophysically measured visual deficits.

Using this strategy, we seek to uncover the relationship between the neural changes that underlie amblyopia and the perceptual consequences of the disorder. We are currently pursuing the idea that the relatively mild type of amblyopia typically produced by anisometropia (having unsymmetric parts) involves a deficit in the P system, while the more complex syndrome that often follows strabismus also involves important deficits in the M system.

Human Retroviral Gene Expression and Cellular Transcription

Gary J. Nabel, M.D., Ph.D.—Associate Investigator

Dr. Nabel is also Associate Professor of Internal Medicine and Biological Chemistry at the University of Michigan Medical School. He received his bachelor's degree from Harvard College, his M.D. degree from Harvard Medical School, and his Ph.D. degree in cell and developmental biology from Harvard University. He was a research fellow at the Whitehead Institute, Massachusetts Institute of Technology, in the laboratory of David Baltimore, before moving to the University of Michigan.

T lymphocytes protect the body from invasion by foreign organisms but can also serve as targets of infection by viruses. One such agent is the human immunodeficiency virus (HIV), which causes the acquired immune deficiency syndrome (AIDS). Under normal circumstances, T cells become activated in response to infection and begin to synthesize a set of proteins that activate the immunologic defense system. In T cells that contain the AIDS virus, cellular activation provides a signal to HIV to stimulate viral replication. Our laboratory has characterized regulatory proteins that stimulate T cell and retroviral gene expression. These cells provide a model to study coordinate gene expression during development and following viral infection.

We have identified proteins that, in binding to control regions, regulate the expression of HIV and other immune system proteins. At the same time, we have begun to use our knowledge of cellular and viral transcription to deliver recombinant genes *in vivo*. This system has allowed us to learn more about the biology of these genes, particularly in endothelial and vascular smooth muscle cells of the vessel wall, and has provided new opportunities for studies on gene transfer.

Regulation of HIV Gene Expression in T Cells and Monocytes

The expression of HIV can be activated in T cells treated with phorbol esters or other immune system activators. We have previously shown that stimulation of these cells increases the binding activity of a protein that binds to a DNA control region, called NF-κB (nuclear factor that recognizes a sequence in the κ immunoglobulin light chain of B cells). NF-κB is responsible for stimulation of HIV transcription in activated T cells. The DNA sequence recognized by this transcription factor is twice repeated in the HIV control region, and mutation of these sites abolishes inducibility of HIV. NF-κB acts in synergy with HIV products, such as the *tat*-I gene, to enhance HIV gene expression in an infected cell.

The transcription factor NF-κB is a protein complex composed of a DNA-binding subunit and an associated transactivation protein (of relative molecular masses 50 and 65 kDa, respectively). Both subunits have similarity with the *rel* oncogene and the *Drosophila* maternal-effect gene *dorsal*. The 50-kDa DNA-binding subunit was previously thought to be a unique protein, derived from the 105-kDa gene product (p105). We have recently reported the isolation of a cDNA that encodes an alternative DNA-binding subunit of NF-κB. It is more similar to p105 NF-κB than other family members and defines a new subset of *rel*-related genes. It is synthesized as a protein of about 100 kDa (p100) that is expressed in different cell types, contains cell cycle motifs, and like p105, must be processed to generate a 50-kDa form.

A 49-kDa product (p49) can be generated independently from an alternatively spliced transcript. It has specific κB DNA-binding activity and can form heterodimers with other *rel* proteins. In contrast to the 50-kDa protein derived from p105, p49 acts in synergy with p65 to stimulate the HIV enhancer in transiently transfected Jurkat cells. Thus p49/p100 NF-κB could be important in the regulation of HIV and other κB-containing genes. The above studies are supported in part by grants from the National Institutes of Health.

In addition to HIV type 1 (HIV-1), a second related virus, HIV type 2 (HIV-2), can induce AIDS. HIV-2, a distinct retrovirus, shares nucleic acid and protein similarity with HIV-1. First described in West Africa, HIV-2 has begun to appear throughout the world. Although HIV-1 and -2 both cause AIDS, the length of the asymptomatic period following infection differs for the two viruses.

Because increased viral replication is associated with progression of HIV-related disease, the rate of disease progression may be influenced by host cell regulatory proteins that activate virus replication. Such proteins could be regulated by distinct cofactors that selectively stimulate cellular activation pathways. These T cell activation pathways regulate specific transcription factors that may contribute to the regulation of the latent phase of HIV infection.

We recently defined the transcriptional regulation and induction of these retroviruses and found that their regulation differs. A distinct T cell activation pathway—triggering of the CD3 component of the T cell antigen receptor complex—stimulates HIV-2 gene expression but does not affect HIV-1. The response to T cell receptor stimulation in HIV-2 is mediated by an upstream regulatory element, CD3R, which a sequence-specific DNA-binding protein of the ets family recognizes.

In addition, at least three other cis-acting regulatory sequences contribute to HIV-2 gene expression, including κB, another ets binding site, and an associated element. Jurkat T leukemia cell lines containing HIV-2 provirus also show increased viral replication following stimulation of the T cell receptor complex, in contrast to HIV-1. These findings suggest that HIV-2 and HIV-1 differ in their transcriptional regulation and induction. These studies also raise the possibility that different cofactors contribute to the activation of AIDS associated with HIV-1 and HIV-2.

Expression of Cellular and Retroviral Vector Genes *in Vivo*

Despite recent advances in the understanding of eukaryotic gene regulation, a major obstacle to the therapeutic management of human disease remains the site-specific expression of genes *in vivo*. Using our knowledge of retroviral gene expression, we have developed systems that utilize viral vectors to express biologically active proteins in cells and tissues *in vivo*. We have devised methods that allow a recombinant gene to be expressed efficiently at a specific site *in vivo* by direct introduction of genetic material at the time of catheterization. A recombinant β-galactosidase gene was expressed in a specific arterial segment *in vivo* by direct infection with a retroviral vector or by DNA transfection using liposomes. Several cell types in the vessel wall have now been transduced with recombinant genes, including endothelial and vascular smooth muscle cells. Site-specific gene expression can therefore be achieved by direct gene transfer *in vivo* and could be applied to the treatment of such human diseases as atherosclerosis, cancer, or AIDS. These studies have been supported by grants from the National Institutes of Health and the American Heart Association.

Biologically active proteins are now being introduced into cells, including disparate histocompatibility antigens, growth factors, growth inhibitors, or immune system proteins. The goal of this research is not only to understand basic mechanisms of gene regulation, transcriptional activation, and viral gene expression, but also to define the biological significance of factors that regulate gene expression in complex organisms and to develop novel molecular interventions for human disease.

Molecular Analysis of Muscle Development and Function

Bernardo Nadal-Ginard, M.D., Ph.D.—Investigator

Dr. Nadal-Ginard is also the Alexander S. Nadas Professor of Pediatrics and Professor of Cellular and Molecular Physiology at Harvard Medical School and Cardiologist-in-Chief at the Children's Hospital, Boston. He received his M.D. degree from the University of Barcelona, Spain, and his Ph.D. degree in biology from Yale University. After training in internal medicine and cardiology, he was a student and postdoctoral fellow with Clement Markert at Yale. He was a professor of cell biology at Albert Einstein College of Medicine before assuming his present position.

OUR laboratory continues to be interested in elucidating the molecular mechanisms that regulate the production and function of the contractile system in muscle cells. This apparatus converts the chemical energy generated by foodstuffs and stored in the form of ATP, and is thus the molecular motor for locomotion and for the heartbeat. One of its fundamental roles is in maintaining cell shape. There are variations of the contractile system in every cell of the organism.

The functional unit of the contractile system is the sarcomere, which is composed of a precisely arranged set of proteins. These are produced by a small family of genes, each making a different isoform. In addition, a single gene can in many cases produce several kinds of proteins by a process called alternative splicing. Combinations of the different isoforms arising from these two mechanisms can lead to millions of different types of sarcomeres. Thus significant functional differences among sarcomeres are produced by changing their components, either through gene switching at the transcriptional level or alternative splicing from the same gene. In addition, the performance of a given sarcomere can be affected by changing the availability of ions in the muscle cells. These three aspects of muscle cell biology continue to be the focus of our research.

Transcriptional Regulation of Contractile Protein Genes

Which sarcomeres assemble in a cell depends on which member of the multigene family of contractile protein genes is expressed at that particular time. To analyze the mechanisms involved in switching from one gene to another in the same gene family, we have concentrated on those coding for the myosin heavy chain (MHC). This is the most important protein of the sarcomere, since it contains the enzymatic activity responsible for generating force. We are currently exploring two main questions in the regulation of the MHC genes: What determines which one is expressed in a given cell type at a particular developmental stage? What determines the level of expression?

To date, the best-characterized muscle-specific regulatory factors are the myogenic basic helix-loop-helix (bHLH) proteins of the MyoD family. Muscle-specific induction by these proteins depends on specific DNA binding to a particular sequence, called an E box, present in many muscle enhancers and promoters. The MyoDs interact with the DNA in conjunction with other HLH proteins that are distributed more widely and are present in limiting amounts.

During the past year we have cloned and characterized two new proteins that function as heterodimeric partners of MyoD. These are alternatively spliced products of gene *E2.2* and are particularly abundant in skeletal muscle, heart, and brain. Several lines of evidence indicate that *in vivo* these proteins are the partners of MyoD in muscle and that they also play an important role in other tissues. The fact that these proteins interact with the product of the retinoblastoma gene suggests that they are involved in the regulation of cell proliferation.

One of the paradoxes of the MyoD paradigm for muscle-specific gene regulation is that not all muscle genes contain E boxes, although they are uniformly required for efficient muscle-specific expression. In addition, many of the genes induced by MyoD in skeletal muscle are also expressed in cardiac and, in some cases, smooth muscle, where myogenic bHLH proteins have not been found. It was expected, therefore, that other transcription factors, which in skeletal muscle might be regulated directly or indirectly by the MyoD family, mediate activation and high-level expression of these genes. A candidate for such a factor is the myocyte-specific enhancer-binding factor 2 (MEF2), which binds to a specific DNA sequence known to be important for high-level expression in skeletal and cardiac muscle.

During the past year we have cloned and characterized a family of MEF2 factors. In humans these factors are encoded by a family of at least four genes, which through alternative splicing generate a larger number of proteins. All these proteins share domains that are strikingly similar to the DNA-binding and dimerization domains of

a recently identified MADS gene family, which includes yeast transcription factors, plant homeotic genes involved in flower morphogenesis, and the serum response factor.

MEF2 mRNAs accumulate preferentially in skeletal muscle, heart, and brain, but post-transcriptional mechanisms must also contribute to the MEF2 tissue specificity. Remarkably, cardiac and smooth as well as skeletal muscle contain saturating levels of MEF2 trans-activating factors that are absent in nonmuscle cells. We have shown that MEF2 is induced in skeletal muscle cells by the MyoD gene family, but that this factor, by itself, is insufficient to produce the whole muscle phenotype. Therefore, although MEF2 is induced in skeletal muscle by bHLH proteins, other lineage-determining pathways must lead to MEF2 expression in nonskeletal muscle tissues.

In this context, the finding that both muscle and brain have high-level expression of the same isoforms strongly suggests that MEF2 also has an important role in neuronal gene regulation. Whether bHLH or other factors induce MEF2 genes in cardiac and smooth muscle, as well as nerve, the regulatory sequences of these genes will serve as powerful tools for the identification of the lineage-determining pathways in these cell types.

Regulation of Protein Diversity by Alternative Splicing

One of the fundamental aspects of pre-mRNA splicing that is particularly important for the generation of different proteins from a single gene is the mechanism by which splice sites are identified. We have previously suggested a scanning model for the location of the 3′ splice site of mammalian introns. We proposed that this site is located by a scanning mechanism searching for the first AG downstream of the branch point/polypyrimidine tract. During the past year we have further confirmed and extended this model.

Recent experiments show that scanning for the 3′ splice site starts at the branch point, not the pyrimidine tract, and proceeds until an AG is recognized. Failure to recognize the most proximal AG can arise from extreme proximity to the branch point, or the AG can be sequestered within a hairpin loop. Once the AG has been encountered, scanning stops, but the spliceosome can still see a stretch of about 12–35 nucleotides downstream. The most competitive AG within this scanning window is then selected as the 3′ splice site. The strongest determinant of competitiveness remains the proximity to the branch point. Thus, in many respects, the scanning model for the 3′ splice site closely resembles the model for translation initiation, both in its simplest formulation and in the predictable exceptions to the general rule.

To analyze the mechanisms involved in alternative splicing, we have continued to focus on the α-tropomyosin (TM) gene. This gene generates a minimum of 10 different isoforms that are tissue specific and developmentally regulated. We have concentrated on the mutually exclusive exons 2 and 3 to elucidate the elements involved in this type of regulation.

In the past we identified the mechanisms involved in the production of the default pattern, which results in the exclusion of exon 2 and inclusion of exon 3 in all cell types but smooth muscle cells. During the past year, we have concentrated on the production of the regulated pattern in smooth muscle cells that results in the inclusion of exon 2 and exclusion of exon 3. This pattern is the result of negative regulation by inhibition of the default pattern and is attributable to two well-defined sequences located 5′ and 3′ from exon 3. These sequences are the binding site for protein factors present in smooth muscle but not in HeLa cells. Experiments are in progress to clone this factor.

The search for splicing factors that interact with the polypyrimidine tract has resulted in the cloning and characterization of a novel essential splicing factor. This is a protein of about 100 kDa that is closely associated with the polypyrimidine tract–binding protein previously described by our laboratory and others. Experiments are in progress to determine the particular stage of spliceosome assembly in which this factor is involved.

The Genomic Response to Growth Factors

Daniel Nathans, M.D.—Senior Investigator

Dr. Nathans is also University Professor of Molecular Biology and Genetics at the Johns Hopkins University School of Medicine. He received his B.S. degree in chemistry from the University of Delaware and his M.D. degree from Washington University. His postdoctoral research was done at the National Cancer Institute and the Rockefeller University. Dr. Nathans is a member of the National Academy of Sciences and serves on the President's Council of Advisers on Science and Technology. He received the Nobel Prize in medicine or physiology in 1978 for the application of restriction enzymes to problems in molecular genetics.

THE proliferation of mammalian cells is regulated by extracellular proteins called growth factors. When a growth factor interacts with its specific cell surface receptor, a cascade of biochemical reactions leads to the sequential activation of specific genes. Research in my laboratory concerns the analysis of this induced genetic program in a mouse fibroblast cell line.

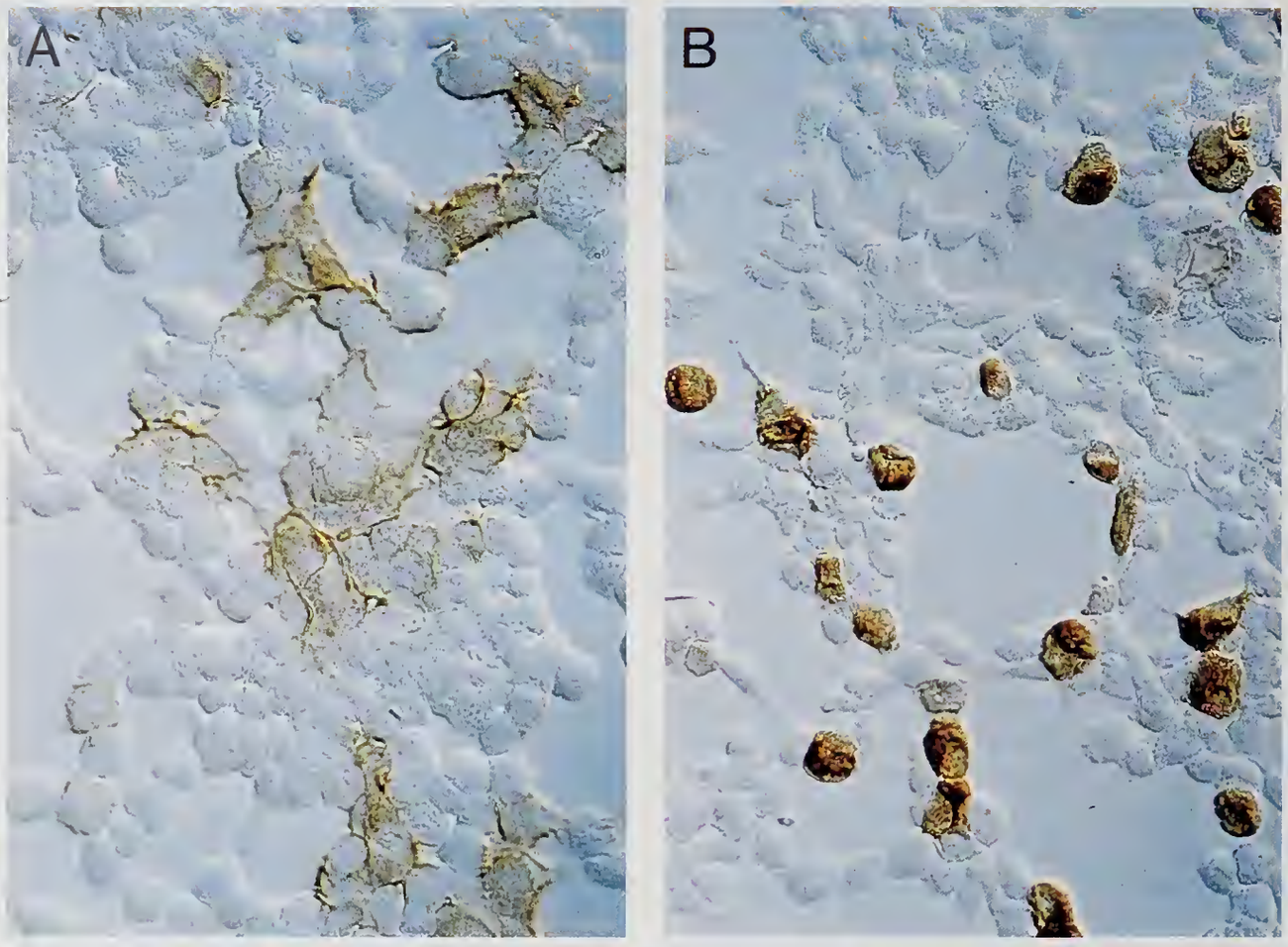
Among the genes expressed in the first wave of gene activation induced by growth factors are many that encode transcriptional regulatory proteins. We have previously described several of these "immediate-early" transcription factors, including members of the Jun and Fos families, zinc finger proteins, and a helix-loop-helix protein. During the past year we have continued our characterization of these proteins and their genes.

Several of the immediate-early transcription factor genes are activated by platelet-derived growth factor (PDGF) and other ligands through one or more serum response elements (SREs) upstream of each gene. In the case of *jun-B*, the upstream sequence near the start of the gene has no functional SRE, nor other signals essential for activation of the gene. By analyzing a series of mutants, a graduate student in the laboratory, Evelio Perez-Albuerne, has found that *jun-B* has regulatory elements downstream of the gene, including a functional SRE and a cyclic AMP response element that mediates the activation of *jun-B* by agents that elevate intracellular cAMP. Thus the mechanism of activation of *jun-B* by serum or PDGF appears to be similar to that previously found for a number of other immediate-early genes, except that the response elements are farther away from the start of transcription and downstream of the gene.

The Jun and Fos family of proteins are DNA-binding transcription factors that form dimers through interacting domains called leucine zippers. Pierre Chevray, another graduate student, has initiated a search for other proteins that interact with Jun and Fos and regulate their activities. For this purpose he used a previously described yeast genetic system that allows one to detect protein-protein interactions and to clone the gene for an interacting protein. By this means he has identified several proteins that interact with the Jun segment that forms a leucine zipper. Among these proteins are another leucine zipper transcription factor, two previously unidentified proteins, and the cytoskeletal proteins α- and β-tropomyosin, which are known to form leucine zippers. We are now exploring the physiological significance of these interactions.

After the appearance of immediate-early transcription factors in growth factor–treated cells, another set of genes comes into play. Activation of these "delayed-early" genes is thought to be mediated by immediate-early transcription factors. Associates Anthony Lanahan and John Williams have cloned and characterized a number of cDNAs corresponding to delayed-early genes. Some of the genes are induced by mitogens in nonfibroblastic cell lines also. Among the proteins they encode are a chromosomal protein, a transmembrane channel protein, an enzyme involved in adenine nucleotide biosynthesis, a protein related to a known cytokine, and several previously unidentified proteins. We are further characterizing some of these. In addition, we are studying the role of immediate-early transcription factors in activating their genes.

Research on the Jun proteins is supported by a grant from the National Institutes of Health.



The commonest rhodopsin mutation among patients with autosomal dominant retinitis pigmentosa has a substitution of the amino acid histidine for proline 23. Antibody-stained human opsin (the apoprotein of rhodopsin), expressed by transient transfection in tissue culture cells, is shown here. A: Wild-type opsin accumulates in the plasma membrane. B: The mutant opsin is predominantly in the endoplasmic reticulum.

From Sung, C.-H., Schneider, B.G., Agarwal, N., Papermaster, D.S., and Nathans, J. 1991. Proc Natl Acad Sci USA *88:8840–8844.*

Molecular Biology of Visual Pigments

Jeremy Nathans, M.D., Ph.D.—Associate Investigator

Dr. Nathans is also Associate Professor of Molecular Biology and Genetics and of Neuroscience at the Johns Hopkins University School of Medicine. His undergraduate work was in biology and chemistry at the Massachusetts Institute of Technology. He received a Ph.D. degree in biochemistry and later his M.D. degree at Stanford University. Before joining the staff at Johns Hopkins, Dr. Nathans spent a year as a postdoctoral fellow at Genentech.

VISUAL pigments are the light-absorbing proteins that initiate phototransduction. Each consists of a chromophore, 11-*cis* retinal, joined to an integral membrane protein, opsin. The visual pigments constitute one branch of a large family of cell surface receptors that transduce external stimuli by activating G proteins. In the visual system, the activated G protein stimulates a cGMP phosphodiesterase, and the resulting transient decline in cGMP closes plasma membrane cation channels.

Photon absorption by 11-*cis* retinal causes it to isomerize from 11-*cis* to all-*trans*. The attached protein then undergoes a series of conformational changes, leading ultimately to a form that interacts with the G protein. The changes underlying visual pigment activation are likely to resemble those that accompany hormone-receptor binding among the other members of this receptor family.

Our laboratory is interested in three general areas related to the visual pigments: their structure and function, the control of their expression, and their variation within the human population.

Structure/Function Studies

Several years ago we developed a system for producing large quantities of bovine rhodopsin by expression of cloned cDNA in tissue culture cells. As described below, we are using this system in conjunction with site-directed mutagenesis to define those residues involved in protein conformational changes, in protein-chromophore interactions, and in protein stability.

Because the retinal chromophore is sensitive to changes in its immediate environment, the various conformations of light-activated rhodopsin each possess distinctive absorption spectra. By following the changes in spectral absorbance following photoactivation, one can determine the quantity and rates of formation and decay of each conformational intermediate. Charles Weitz, a postdoctoral fellow, has used this assay to map those amino acids that play an important role in rhodopsin's transition to the active conformation, i.e., the conformation that interacts with the G protein.

In one set of experiments, we sought to examine the mechanism responsible for the pH dependence of this transition. Thirty years ago George Wald and his colleagues observed that low pH favors this transition and that the pH dependence was consistent with a mechanism in which protonation of one or more histidines was tightly coupled to the transition. We therefore mutated one-by-one each of the six histidine residues to phenylalanine and monitored the ability of the mutant rhodopsin to form the active conformation. All of the mutant proteins could bind to 11-*cis* retinal and form a light-sensitive pigment, but mutants in which histidine^{211} was replaced either by phenylalanine or cysteine were unable to assume the active conformation.

The simplest interpretation of this experiment is that the histidine^{211} is the site at which protonation drives rhodopsin into its active conformation. Dr. Weitz has pursued this observation by identifying mutants at other sites that have the converse property: they form the active conformation more efficiently than the normal protein. This second type of mutant therefore identifies amino acids that normally act to keep the protein in the inactive conformation. As most amino acids have little or no effect on this conformational transition, it should be possible to identify the handful of critical residues that control it.

In the human retina, rhodopsin mediates vision in dim light, whereas a related set of visual pigments, the cone pigments, mediate bright light vision as well as color vision. The spectral properties of the cone pigments have been of great interest to physiologists and psychologists and have been the object of investigation for over a century. Unfortunately, the instability and scarcity of these proteins make them difficult to study. Several years ago we isolated the genes that encode the human cone pigments. Using these reagents, Shannath Merbs, a graduate student, has recently succeeded in producing large quantities of the pigments and determining their precise absorption spectra.

Interestingly, the human gene pool contains two versions of the red pigment. One carries an alanine at position 180 and absorbs maximally at 552 nm, whereas the second version carries a serine at position 180 and absorbs maximally at 557 nm. This spectral difference explains a number of long-standing observations regarding differences in color vision among otherwise color-normal individuals. We are now using this system to study hybrid pigments encoded by hybrid genes that were generated by recombination between the red and green pigment genes. Hybrid genes of this type are carried by 7 percent of human X chromosomes and account for most of the common forms of red-green color blindness.

Control of Visual Pigment Gene Expression

We have been interested for some time in the general question of how different cells in the retina assume their correct identities. As an initial approach, we have examined the control of visual pigment gene expression. Each photoreceptor cell appears to produce only a single type of visual pigment: rhodopsin in the rods, and the red, green, and blue pigments in their respective cone types. As a working model, we assume that this specificity in protein production reflects a corresponding specificity at the level of gene transcription.

One region of DNA that is important for activation of the red and green pigment genes has recently been identified. Over the past several years, we have studied families with a rare X-linked disorder called blue cone monochromacy, in which both red and green cone sensitivities are absent. In many families a DNA deletion is observed in which sequences are removed adjacent to the cluster of red and green pigment genes on the X chromosome. This set of deletions defines a region of 0.6 kilobases that appears to be required for normal red and green visual pigment gene function, even though the start sites of transcription of these genes are, respectively, 3 kilobases and 42 kilobases away.

Yanshu Wang, a graduate student, has recently shown that a segment of human DNA encompassing this essential region and including the red pigment gene promoter directs expression of a linked reporter gene to cone cells in the mouse retina. An otherwise identical construct from which the 0.6-kilobase segment has been removed is inactive. We have observed that within this essential segment, there is a smaller region that has a high degree of DNA sequence homology across species. Most likely, this region represents a binding site for one or more transcription factors present in the red and green cones.

Retinitis Pigmentosa

During the past several years, we have begun to work on a group of inherited retinal diseases called retinitis pigmentosa (RP). The hallmarks of RP are night blindness and a slow progressive loss of peripheral vision, leading in most cases to complete blindness by the fifth or sixth decades of life. RP affects one person in 4,000 in all populations examined. Last year Ching-Hwa Sung, a postdoctoral fellow, reported finding rhodopsin mutations in one-quarter of patients with autosomal dominant RP, an inheritance pattern that is found in approximately one-fifth of the patient population. In a group of 161 unrelated patients, 13 different mutations were identified, and in all cases the mutations co-inherited with the disease in affected families.

Dr. Sung has produced each of the mutant opsins, as well as normal human opsin, in tissue culture cells and has analyzed their biochemical properties. The mutant proteins fall into two classes. Members of one class resemble wild-type opsin in yield, ability to bind 11-*cis* retinal, and efficient transport to the cell surface. By contrast, members of the second class are produced in low yield, bind 11-*cis* retinal variably or not at all, and are transported inefficiently to the cell surface. These characteristics suggest that the second class of mutant proteins are either incorrectly folded or unstable. It seems reasonable to suppose that production of large quantities of an unstable opsin may be deleterious to the photoreceptor. The biochemical defect in the first class of mutant proteins is not apparent from the experiments performed thus far and is currently under investigation.

Gene Regulation in Animal Cells

Joseph R. Nevins, Ph.D.—Investigator

Dr. Nevins is also Professor of Genetics and of Microbiology and Immunology at Duke University Medical Center. He received his Ph.D. degree in virology at Duke University, where he studied with Wolfgang Joklik. His postdoctoral studies as a Jane Coffin Childs fellow focused on the mechanisms of mRNA biogenesis and were conducted with James Darnell at the Rockefeller University, where he later became a faculty member.

REGULATION of gene expression is central in the determination of cellular phenotype and in the complex transformations that take place during such events as oncogenesis. The goal of our laboratory is to elucidate the molecular mechanisms of gene control pathways.

Adenovirus Transcriptional Regulation—A Role in Oncogenesis

Complex cellular events are often best studied through the use of simple model systems. The control of transcription (transfer of DNA-coded information to RNA to guide protein synthesis) mediated by viral regulatory proteins is instructive as regards the workings and control of transcription factor activity in eukaryotic cells.

Work in our laboratory has focused on the elucidation of transcription control by viral proteins, using *in vitro* assays. We have found that a cellular transcription factor termed E2F is regulated by other cellular proteins and that these interactions prevent E2F from being co-opted by the viral genome. The adenovirus regulatory protein E1A, however, can dissociate these complexes, releasing E2F, which can then be utilized by another viral product, the E4 protein. This redirects the cellular transcription factor for viral purposes.

These findings are significant for several reasons. First, they provide insight into the evolution of viral regulatory events that manipulate the host cell for the virus's benefit. In particular, they highlight the interplay between two viral proteins (E1A and E4) that together redirect the E2F factor to a viral-specific role.

Second, this activity of E1A correlates with the oncogenic activity of the protein. The sequences in E1A that are responsible for this activity are shared with other viral oncogene products, such as the SV40 T antigen and the human papillomavirus E7 protein. Each of these viral proteins has been shown to activate transcription, dependent on E2F.

Third, our recent experiments have shown that the viral T antigen and E7 protein can also dissociate complexes containing the E2F transcription factor. In short, this latest discovery has defined a common biochemical activity of these viral regulatory proteins that is likely part of their oncogenic activity.

The identification of the cellular proteins that are complexed to the E2F factor has provided important insight into the role of E1A as an oncogene. We have found that there are multiple such E2F complexes and that their formation is regulated during the cell cycle. One complex contains the cell cycle–regulated protein cyclin A, and another the retinoblastoma gene product.

The product of the retinoblastoma susceptibility gene (*RB1*) is a 110-kDa nuclear phosphoprotein that is expressed at equivalent levels in all cell types examined except certain tumor cells in which the *RB1* gene has been inactivated by mutation or deletion. Given that it is the loss of *RB1* function that is correlated with the development of certain human tumors, it is widely believed that the Rb protein functions to limit or constrain cell proliferation.

It now appears that the ability of E1A, as well as T antigen and E7, to dissociate the E2F-Rb complex, releasing free E2F, is at least one of the events of Rb inactivation. The E2F-Rb complex, like the E2F-cyclin complex, is also regulated by cell proliferation. Moreover, the E2F-Rb complex is absent in cells that express a nonfunctional form of the Rb protein.

This latter result strongly argues that E2F is a functional target for Rb action. Thus the consequence of either E1A action, the deletion or mutation of Rb, or the phosphorylation of Rb is the loss of the E2F-Rb interaction and the generation of free E2F molecules. In short, "inactivation" of Rb can be viewed as the loss of the E2F-Rb interaction. Moreover, it would appear that the consequence of disruption of either the E2F–cyclin A complex or the E2F-Rb complex is the release of transcriptionally active E2F and that this event may be important for stimulating the transcription of genes whose products are critical for enhancing cell proliferation.

Molecular Mechanisms for Polyadenylation

The generation of the 3′ terminus of the mature mRNA, commonly termed poly(A) site formation,

is a critical event in the biogenesis of most mRNA molecules. Moreover, since many transcription units encode multiple mRNAs that utilize alternative poly(A) sites, there is a potential regulatory role for polyadenylation. Indeed, analysis of the formation of the immunoglobulin heavy-chain transcripts during B cell differentiation suggests that alternative poly(A) site choice is a contributing factor in the switch from production of the Ig mRNA that encodes the membrane-bound form of the protein to the mRNA encoding the secreted form of the protein. Thus an understanding of the mechanism of poly(A) site utilization, including the sequences directing the processing event and the factors involved in processing, is an important focus of our work.

We have recently used a biochemical approach to investigate the molecular mechanism of polyadenylation. Factors have been purified from HeLa cell nuclear extracts that can reconstitute accurate and sequence-specific poly(A) site processing *in vitro*. Two of these factors interact specifically with the pre-mRNA, and an analysis of this interaction suggests a pathway of assembly of a functional poly(A) site-processing complex.

The interaction of one factor with the conserved AAUAAA element is specific, but the RNA-protein complex that forms is very unstable and rapidly dissociates. The interaction of a second factor with the RNA requires a distinct sequence element but also requires the prior interaction of the first factor with the AAUAAA element. Most importantly, the resulting ternary complex is stable and does not readily dissociate.

These observations suggest a pathway of assembly of poly(A) site factors that involves an initial recognition of the AAUAAA element followed by a commitment step in which the second factor stabilizes the overall complex. Since the relative stability of this complex reflects the relative efficiency of processing, it appears that these interactions may be important regulatory targets for the control of polyadenylation.

This work on the mechanisms of polyadenylation is supported by a grant from the National Institutes of Health.

Molecular Genetics of X-linked Disease

Robert L. Nussbaum, M.D.—Associate Investigator

Dr. Nussbaum is also Associate Professor of Human Genetics, Pediatrics, and Medicine at the University of Pennsylvania School of Medicine and Consultant in Clinical Genetics at the Children's Hospital of Philadelphia. He received his undergraduate training in applied mathematics at Harvard College and his M.D. degree at Harvard Medical School in the Harvard-MIT Joint Program in Health Sciences and Technology. After his residency in internal medicine at Barnes Hospital, Washington University School of Medicine, he moved to Baylor College of Medicine, first for a genetics fellowship with Thomas Caskey and Arthur Beaudet and later as a faculty member. He then moved to the University of Pennsylvania, where he developed his research program in molecular genetics and its application to the diagnosis and elucidation of human genetic disease.

THE research in my laboratory is directed toward elucidating the molecular bases for a number of human genetic diseases. Each disease under investigation is known to be caused by a gene on the X chromosome, but the molecular mechanism, the gene involved, and the nature of the underlying mutations have been hitherto unknown. Recombinant DNA techniques have been employed to isolate the responsible genes, with the aim of furthering our understanding of the normal processes that when disrupted result in each of these diseases.

Choroideremia

Choroideremia is a rare X-linked disease of the retina that produces blindness in affected males. The gene responsible and the mechanism of retinal damage have until recently been unknown. Our laboratory is using information about where the choroideremia gene is to identify it and explain why mutations in this gene cause the disease.

We have isolated a gene from the region around a chromosomal translocation in a female patient with choroideremia. Her disease resulted from disruption of the choroideremia gene by the chromosomal break in the X chromosome in this region. A transcribed gene that is disrupted by this chromosomal translocation has been identified and found to be very similar, although not identical, to one isolated in the laboratory of Frans Cremers by his study of males with choroideremia and submicroscopic deletions. The expression of this gene, at the level of mRNA, is abnormal in 75 percent of patients with choroideremia.

The predicted protein sequence of the gene identified in both laboratories has a subtle similarity with a protein isolated from platelets that may be involved in regulation of the G proteins, a very large and heterogeneous group involved in carrying certain intracellular signals. The cellular mechanism by which vision occurs, known as visual transduction, is the entire pathway by which light, the initial signal, is translated first into biochemical reactions and then nervous impulses in the retina. A growing body of evidence suggests that abnormalities in the proteins that carry out visual transduction contribute to a variety of degenerative retinal diseases that lead to blindness.

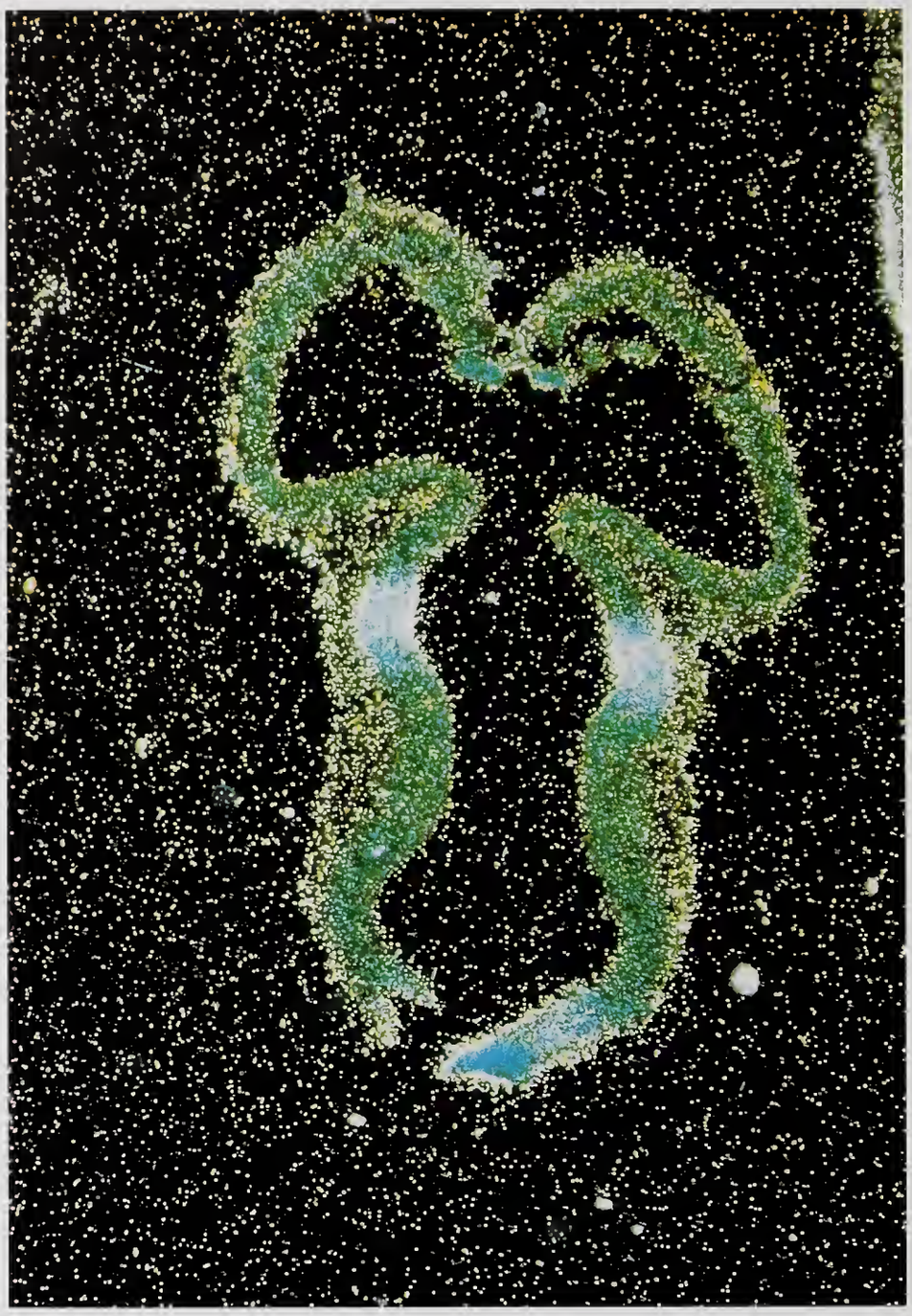
The plan is to investigate the exact role of the choroideremia gene product in visual transduction, in an attempt to understand how abnormalities in that gene lead to retinal degeneration. This work is supported by a grant from the National Institutes of Health.

Lowe's Syndrome

Lowe's syndrome is an uncommon X-linked disease that causes mental retardation, cataracts, and kidney dysfunction. The cause is unknown. As with choroideremia, our strategy is to identify the Lowe's syndrome gene through information about its location.

We have isolated two cDNA sequences from the region around a translocation breakpoint in a female with Lowe's syndrome. The transcript detected by one of these cDNAs is disrupted by the translocation and is absent or abnormal in nine unrelated males with the disease. This gene encodes a protein involved in the metabolism of inositol phosphates, a very heterogeneous but important class of compounds with a complex biochemistry. These structural components of normal cell membranes also play an important role in anchoring proteins to cell membranes and in intracellular signaling pathways.

Our goal is to understand the biochemical processes that, when defective, lead to brain, lens, and kidney dysfunction and damage. Insights into normal lens formation and normal brain and kidney function could result, and methods for improved diagnosis and therapy for the disease may be found. This work is supported by a grant from the National Institutes of Health.



Expression pattern of the Wnt-3 *gene in the developing mouse brain.* In situ *hybridization at 10 days of gestation shows* Wnt-3 *expression in one of the compartments of the diencephalon, D2, and in the developing midbrain-hindbrain junction that give rise respectively to the dorsal thalamus and the cerebellum.*

Research and photograph by Patricia Salinas in the laboratory of Roel Nusse.

Function of Oncogenes in Early Embryogenesis

Roel Nusse, Ph.D.—Associate Investigator

Dr. Nusse is also Associate Professor of Developmental Biology at the Stanford University School of Medicine. He obtained his Ph.D. degree from the University of Amsterdam and was a postdoctoral fellow with Harold Varmus at the University of California, San Francisco, before returning to Amsterdam, where he became head of the Department of Molecular Biology at the Netherlands Cancer Institute. Several years ago he moved to the Beckman Center of Stanford University and the Howard Hughes Medical Institute. Dr. Nusse is a member of the European Molecular Biology Organization (EMBO).

IT is now well established that genes whose altered expression can lead to cancer (generally called oncogenes or proto-oncogenes) are indispensable regulators of normal cell proliferations. Many oncogenes participate in the cells' decisions whether to divide or remain quiescent. Apart from the regulation of cell proliferation, growth in a normal organism obviously needs control at another level: during the formation of patterns in which cells become properly arranged. Over the past few years, it has become increasingly clear that oncogenes are among the key regulators of this aspect of growth control also.

The evidence for such controlling functions has primarily come from the genetic dissection of morphogenesis in organisms such as *Drosophila* and *Caenorhabditis elegans*. Many of the developmental genes isolated from those animals are highly homologous to mammalian oncogenes, in particular those that encode signaling molecules involved in cell-to-cell communications during embryogenesis. Now, with the increasing possibilities to identify developmental genes in the mouse as well, additional examples have emerged for a link between cancerous growth and the control of normal development.

The *Wnt/wingless* gene family is a paradigm for the connections between tumorigenesis and embryogenesis. The prototypic member of this group, *Wnt*-1, is an oncogene frequently activated in mouse mammary cancer. The *Wnt*-1 gene is normally not expressed in the mammary gland or in most other adult tissues. In tumors, however, its transcription is induced by nearby insertion of proviral DNA of a retrovirus, the mouse mammary tumor virus. Proof that *Wnt*-1 is an oncogene came from transfection experiments and from the finding that introduction of *Wnt*-1 as a transgene into the germline of mice can lead to tumor induction. *Wnt*-1 is part of a family of genes that in the mouse consists of at least 10 members. All of these genes encode secreted proteins rich in cysteine residues.

Over the past few years, evidence that the *Wnt* genes are prime determinants for early development has accumulated from many different corners. For example, most of these *Wnt* genes have a very restricted pattern of expression during early developmental stages, in organisms ranging from mice to fruit flies. More telling are the findings that *Wnt* gene mutations prevent normal development of the mouse brain and normal segmentation of *Drosophila* embryos. Moreover, it has been shown that ectopic expression of *Wnt* genes induces axis duplication in frog embryos.

The work in our group is focused on the role of several members of the *Wnt* gene family in the development of the mouse, along with investigations of *Wnt* genes in the fruit fly. In the mouse we perform detailed *in situ* RNA hybridization analysis of the expression of *Wnt*-3 and a highly related gene, *Wnt*-3A. Both genes are expressed in the developing neural tube.

In particular, the anterior boundary of expression of *Wnt*-3 and *Wnt*-3A is interesting. The genes are expressed in the diencephalon and in the cerebral hemispheres, suggesting that they play important roles in establishing these compartments in the developing brain. For example, *Wnt*-3 is expressed in the D2 neuromere of the developing diencephalon during day 9 and 12 of embryogenesis. The expression precedes the subdivision of the diencephalon into the ventral and dorsal thalamus (see figure), suggesting a functional role for *Wnt*-3 in this process. Another remarkably restricted expression pattern of *Wnt*-3 is seen in the cerebellum, where the gene is exclusively expressed in the Purkinje cells.

Understanding the mechanism of action of the *Wnt*-1 gene family during embryogenesis has been complicated by difficulties in characterizing the *Wnt* proteins. We wish, for example, to identify the receptors for these secreted molecules, but as no biologically active *Wnt* gene product is available, the nature of the receptors remains elusive. In parallel to these biochemical experiments, we study the interactions between *Wnt* and other genes by taking advantage of the extensive genetic analysis of *Drosophila* embryogenesis. Some years ago we made the surprising finding that the *Drosophila Wnt*-1 homologue is

identical to the segment polarity gene *wingless*. This observation has allowed us to study the interactions of *Wnt*-1/*wingless* with other segmentation genes.

The basic body plan of the fruit fly is established by several classes of genes that progressively divide the embryo into smaller compartments: the gap genes, the pair-rule genes, and the segment polarity genes. Whereas the gap and pair-rule genes encode nuclear proteins and are active before the *Drosophila* embryo becomes cellularized, the segment polarity genes are the first that control cell-cell interactions. The *wingless* gene is a good example, encoding a secreted factor, but other segment polarity genes are thought to interact with *wingless*. To study the properties of the *wingless* protein, we have made antibodies that recognize it in whole-mount embryos and in individual cells. The protein is seen on the surface of cells and in intracellular vesicular structures. We have characterized the distribution of the *wingless* protein in embryos that are mutant for *wingless* itself and have obtained evidence that lack of secretion of the protein is the primary defect in these mutants.

In an additional approach, we have overexpressed the *wingless* gene in embryos, from a heat-shock promoter. The phenotype of the heat-shock *wingless* embryos is the opposite of the phenotype of embryos lacking *wingless*. Surprisingly, heat-shock *wingless* embryos are very similar to embryos that are mutant for two other segment polarity genes, called *naked* and *zeste-white-3*, suggesting common biochemical pathways of *wingless* and those genes.

We have also found that *wingless* in *Drosophila* is part of a gene family, with at least two additional members. These genes, called D*Wnt*-2 and D*Wnt*-3 for the time being, are also expressed during early embryogenesis, in characteristic patterns that differ from *wingless*. Mutants at these genes have not yet been obtained, but the unusual character of the D*Wnt*-3 protein, which is three times as long as other *Wnt* proteins, may give us new clues as to the biochemical mechanism of action of the genes.

In the future we hope to extrapolate our increasing understanding of the function of the *Wnt* genes in embryogenesis to the mechanism of action of the genes in cancer.

Molecular Mechanisms That Regulate B Cell Development

Michel C. Nussenzweig, M.D., Ph.D.—Assistant Investigator

Dr. Nussenzweig is also Assistant Professor and Head of Laboratory at the Rockefeller University. He received his undergraduate and M.D. degrees from New York University and his Ph.D. degree from Rockefeller. He completed his residency and clinical fellowship at Massachusetts General Hospital and conducted postdoctoral research at Harvard Medical School with Philip Leder.

WE are interested in understanding the molecular mechanisms that regulate B lymphocyte activation and differentiation. Our approach has been to focus on one important transition in B lymphocytes, allelic exclusion.

The immune system is responsible for protecting vertebrates from both invasion by infectious organisms and deregulated growth of endogenous malignancies. To accomplish this task, the immune system must be able to distinguish self from non-self. Evolution has solved this problem in higher vertebrates by providing a network of cell types and humoral agents. It is the lymphocytes—T cells and B cells—that direct the specificity of immune responses.

Although the mechanism of antigenic recognition differs for these two cell types, the generation of receptor diversity is achieved in a similar fashion. In both cases the business end of the receptor is created in individual somatic cells by a series of genetic recombinations at a minimum of two loci. For example, immunoglobulin molecules that serve as the B cell receptor are composed of two sets of rearranging genes that encode the heavy- and light-chain immunoglobulin proteins. Furthermore, the same genes that encode the B lymphocyte receptor also direct the production of secreted antibodies that are an important component of humoral immune responses. Thus the regulation of T cell and B cell receptor rearrangements is a central feature of the generation of immune responses.

The joining events that bring together the immunoglobulin segments occur in an ordered and regulated fashion. In B lymphocytes, rearrangements begin at the heavy-chain locus with the recombination of D and J segments. This is followed by joining of DJ with one of 100–1,000 variable-region segments. After a functional immunoglobulin heavy-chain transcription unit is created, the light-chain genes undergo a similar set of rearrangements.

One poorly understood aspect of these events is the ability of lymphocytes to limit themselves to the production of a single receptor. Since productive rearrangements could occur in two heavy-chain and four light-chain alleles, a single B lymphocyte could potentially make several types of receptors, including hybrid molecules. The mechanism that ensures that only one receptor is produced is referred to as allelic exclusion. It is an important safeguard for the immune system, since production of multiple receptors by a lymphocyte would dilute the specificity of any given immune response.

Much of the early work in the area of allelic exclusion was based on examining the status of immunoglobulin genes in transformed B cells. The transformed cells are frozen in one stage of lymphocyte development and for this reason offer only a static picture of important regulatory events. Unfortunately there is no *in vitro* system that faithfully reproduces regulated immunoglobulin gene rearrangements. To study how immunoglobulin genes can regulate allelic exclusion, we turned to transgenic mice.

Our approach has been to introduce into the germline of mice human immunoglobulin genes that have been modified to direct the synthesis of either membrane-associated or secreted immunoglobulin heavy chains. We found that the expression of human membrane-bound immunoglobulin M (IgM) results in the exclusion of most endogenous mouse immunoglobulins. The secreted version of the same transgene has little effect.

This initial observation raises two important questions. First, how does membrane immunoglobulin signal? Second, how is exclusion regulated at the genetic level? During the past year we have made significant progress toward resolving both questions. To understand signaling by the membrane-associated immunoglobulin molecule, we have developed a system for complete functional reconstitution of the immunoglobulin receptor from cloned components in heterologous cells. Transport of IgM to the surface of T cells requires coexpression of the immunoglobulin heavy and light chains with B29, an immunoglobulin-associated polypeptide. In addition, the transfected receptor is fully active in the presence of B29. MB1, a second IgM-associated poly-

peptide, is not required for transport or signal transduction. Progress in understanding the functional role of the immunoglobulin receptor subunits in signaling has been hindered by the multisubunit nature of the receptor. In addition, transfected immunoglobulin expressed on the surface of fibroblasts does not appear to be functional for signal transduction, even in the presence of B29 and MB1. Our ability to produce a functional receptor by transfection in T cells establishes that there is enough structural similarity between the T and B cell signal transduction pathways to allow recognition of the IgM-B29 complex in T cells. This observation should simplify the structural and functional analysis of the IgM antigen receptor and ultimately allow us to understand how membrane immunoglobulin signals for antigen recognition and allelic exclusion.

In addition to our ability to study the structural requirements for immunoglobulin function, our experiments also have potential clinical implications. One major difference between the T cell receptor and immunoglobulin antigen receptors is the nature of the antigen recognized. Immunoglobulins recognize antigens directly; the major histocompatibility complex (MHC) restricts recognition of antigen by the T cell receptor. The requirement for MHC recognition severely limits the targets recognized by T cells and restricts transfer of cellular immunity. Our reconstitution experiments offer one potential solution to this difficult clinical problem—the production of T cells that utilize immunoglobulins as antigen receptors. In T cells that recognize antigen with immunoglobulin receptors there is no MHC restriction, and thus this barrier to transfer of cellular immunity is potentially abrogated.

Our second goal has been to understand allelic exclusion at the genetic level. We have documented that chromosomal position plays an important role in the regulation of gene rearrangements in the immunoglobulin locus. In addition we have started to investigate the regulatory function of the recombinase genes *RAG-1* and *RAG-2* in allelic exclusion. Deregulating the expression of the *RAG* genes in the lymphocytes of transgenic mice has profound effects on lymphocyte differentiation and function. Animals that carry *RAG-1* and *RAG-2* transgenes develop hepatosplenomegaly and a profound lymphoproliferative disorder that is rapidly lethal. We are pursuing a molecular understanding of the pathogenesis of this disorder.

Mechanism of DNA Replication

Michael E. O'Donnell, Ph.D.—Assistant Investigator
Dr. O'Donnell is also Associate Professor of Microbiology at Cornell University Medical College, New York City. He received his Ph.D. degree from the University of Michigan, Ann Arbor, where his research on electron transfer in the flavoprotein thioredoxin reductase was conducted with Charles H. Williams, Jr. He performed postdoctoral work on Escherichia coli *replication with Arthur Kornberg and then on herpes simplex virus replication with Bob Lehman, both in the Biochemistry Department at Stanford University.*

MY laboratory is studying the duplication of genetic information. By understanding the fundamental mechanisms of cell growth, or the replication of DNA, we may obtain insights into the development of abnormal cells, including tumor cells.

The genetic material of our cells, the chromosomes, is a library with all the information needed for the multitude of duties required to maintain the cell's life. Included in these duties is the buildup of complete, new cellular machinery for the synthesis of another cell (reproduction). The chromosome library is made of two long interwound helical fibers of DNA (deoxyribonucleic acid polymers). Before a cell can divide to form two new cells, it must duplicate the genetic library so that each cell has a complete copy of instructions on how to live.

The process of duplicating DNA is intricate, and the cell has evolved a precision machine to carry out this important task. Its several protein parts are like gears of a machine, which coordinate their actions to unzip and unwind the double-helical strands of DNA. The machinery then uses the separated single strands as templates to synthesize two double-helical daughter chromosomes. Subsequently these will segregate in two newly formed cells.

Our goal is to understand, at a molecular level, the workings of proteins in the mechanics of DNA duplication. The system we are studying is the bacterium *Escherichia coli*, a relatively simple organism. The protein machine that duplicates the *E. coli* chromosome is called DNA polymerase III. The DNA polymerase III of *E. coli* has nine accessory proteins plus the polymerase. Like the *E. coli* polymerase III, the DNA polymerases that replicate the chromosomes of higher organisms such as yeast, *Drosophila*, and humans are also composed of a DNA polymerase protein and several other "accessory" proteins.

The function of the DNA polymerase protein (the α-subunit) is to synthesize the DNA. One of the accessory proteins, the ϵ-subunit, is an exonuclease that "proofreads" the product of the polymerase protein. Very little is known about the functions of the other eight accessory proteins. However, since the several accessory proteins to the DNA polymerase are conserved in evolution from bacteria to humans, it seems reasonable to expect their individual functions to serve very important roles in the process of chromosome duplication. Analysis of the *E. coli* DNA polymerase III system will likely extend and generalize the understanding of the replication process in all organisms.

We have recently developed methods to obtain pure preparations of each protein, or subunit, of the *E. coli* DNA polymerase III, and from these the whole complex can be reassembled. We have studied the individual subunits for biochemical activities and for their physical interactions. Two subunits, γ and δ, bind to each other to form a complex that, upon binding to primed DNA, hydrolyzes ATP. In the presence of the β-subunit, the $\gamma\delta$ heterodimer couples the hydrolysis of ATP to clamp a dimer of the β-subunit onto primed DNA. One molecule of the $\gamma\delta$ heterodimer can clamp many β-dimers onto primed DNA. Our biochemical studies indicated that the β-dimer is clamped to DNA by encircling it like a doughnut. The x-ray structure of this β-clamp has recently been solved in a collaboration with John Kuriyan (HHMI, Rockefeller University). It appears as a thin disk with a hole through the middle to accommodate the DNA. These studies on the β-clamp and the $\gamma\delta$ heterodimer were funded by the National Institutes of Health.

The β-clamp on DNA binds the polymerase subunit, tethering it to the DNA template. Whereas the polymerase alone is slow (20 nucleotides/second), it is greatly accelerated upon binding the β-clamp (700 nucleotides/second) and replicates an entire 8-kb single-strand circular DNA without coming off (processive). This fits nicely with the fact that the *E. coli* cell duplicates its 4 million–base chromosome within 30 minutes.

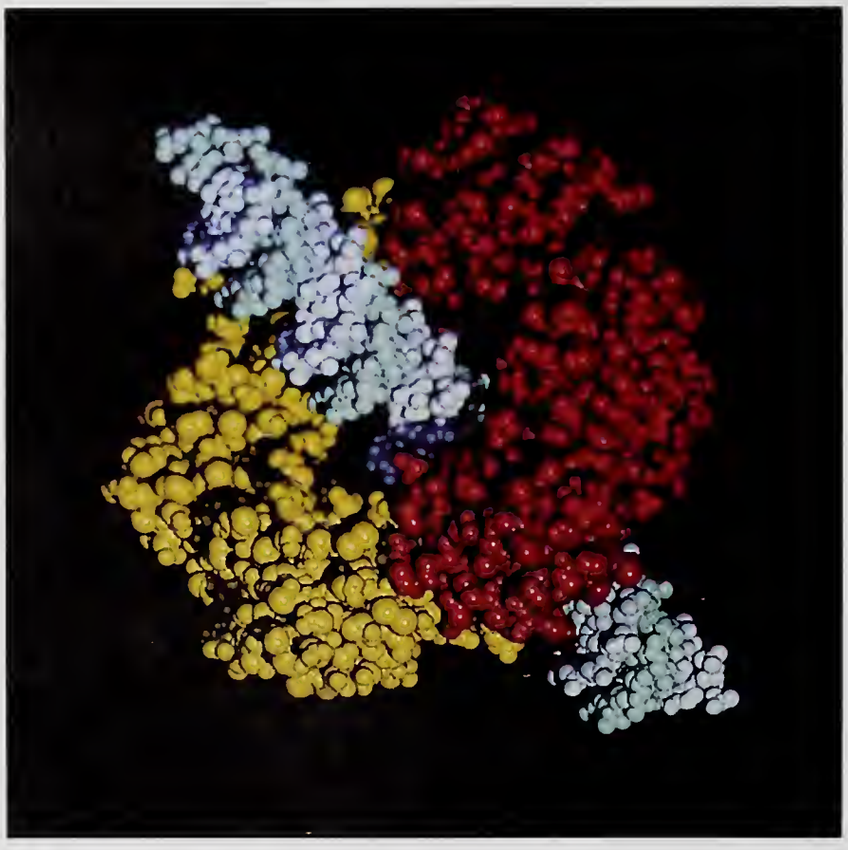
Another function of the γ-, δ-, and β-subunits is to rapidly deliver the polymerase subunit from a completely replicated DNA molecule to a new primed DNA template. This rapid delivery of polymerase is important because one strand of the

DNA duplex (lagging strand), as a result of the geometry of the DNA helix, must be replicated in fragments. Synthesis of these fragments requires that the polymerase subunit be used over and over every 1–2 seconds by repeatedly delivering it from a finished fragment to a new one. Studies on the mechanism of rapid polymerase delivery to new primed templates were supported by a grant from the National Institutes of Health.

Two τ-subunits bind tightly to each other (dimeric), and each binds a polymerase molecule. Hence the τ-subunit dimer serves as a scaffold to form a twin polymerase. Since the chromosome has two strands of DNA, both of which must be replicated, the twin polymerase likely serves the function of coordinately replicating both DNA strands at the same time. These studies on the dimeric polymerase were also supported by a grant from the National Institutes of Health.

These studies on four (γ, δ, β, τ) of the eight accessory proteins of *E. coli* DNA polymerase III have been greatly aided by having the genes (the informational area in the DNA) for three of them (γ, β, τ). This has provided large quantities of proteins for studies via molecular cloning and overproduction techniques. We have recently discovered the genes for each of the remaining five proteins (δ, δ', χ, ψ, θ) of DNA polymerase III and have used these genes to produce and purify large quantities of these proteins. Identification of the genes encoding δ', χ, ψ, and θ was supported by a grant from the National Institutes of Health.

Toward a goal of understanding the overall structure of the holoenzyme, we are using the pure subunits to define the various subunit contacts within the holoenzyme and to assemble the entire replicating machine from its separate parts. Our plans to identify the individual function of each subunit include biochemical analysis and also use of the newly discovered genes for the holoenzyme subunits to construct genetic mutants of *E. coli* for further clues to the function of each subunit. In addition to the polymerase holoenzyme, the helicase and priming proteins are also central to the process of chromosome replication, and we have initiated studies to examine how these proteins coordinate their actions with the holoenzyme.



Model of the β-clamp of the enzyme DNA polymerase III. The function of the clamp is to tether the enzyme to the DNA, greatly accelerating DNA synthesis. The two parts of the β-subunit dimer, represented in red and yellow, encircle the DNA (in green and gray), which has been modeled in the central cavity.

From Kong, X.-P., Onrust, R., O'Donnell, M., and Kuriyan, J. 1992. Cell *69:425–437. Copyright © 1992 by Cell Press.*

Large-Scale Analysis of Yeast and Human DNA

Maynard V. Olson, Ph.D.—Investigator

Dr. Olson is also Professor of Genetics at the Washington University School of Medicine. He was trained as a chemist, receiving his B.S. degree from the California Institute of Technology and his Ph.D. degree from Stanford University. After five years on the faculty at Dartmouth College, he moved to the University of Washington and changed fields from chemistry to genetics. He has served on the National Research Council Committee on the Mapping and Sequencing of the Human Genome and presently serves on the NIH Program Advisory Committee of the National Center for Human Genome Research.

MOST human cells contain 6 billion base pairs of DNA. Embedded therein are an unknown number of genes, perhaps 100,000, that direct the biochemical events in the cells. At present, well over 95 percent of this DNA is unexplored.

Geneticists have developed powerful methods with which to study DNA in small packets. Gene-splicing techniques, DNA sequencing, and methods of reintroducing altered DNA molecules into cells allow the detailed structural and functional analysis of DNA molecules containing up to tens of thousands of base pairs. Our laboratory seeks to extend these approaches to encompass molecules ranging up to millions of base pairs in size.

In the short run, this research should allow the analysis of larger functional units of DNA—for example, large human genes, clusters of coregulated genes, and such structures as centromeres and telomeres, which govern the behavior of human chromosomes during the cell division cycle. In the long run these methods should allow the systematic analysis of the whole human genome —the entire complement of DNA sequences— thereby creating tools, such as detailed maps, that would be of permanent value in biology.

Our approach has been to build on progress in the genetic analysis of microorganisms, particularly yeast. The common laboratory yeast *Saccharomyces cerevisiae*, familiar from its use in baking and wine making, is an ideal model for studies of human cells, because its genetic organization and biochemical pathways are similar to those in higher organisms. It is a powerful tool as well, because human DNA can be altered by gene-splicing techniques into a form that is stably propagated in yeast.

We are just completing a long-term project to map the 14 million base pairs of DNA present on the 16 natural yeast chromosomes. The yeast chromosomes have been mapped at a resolution (i.e., the average spacing between landmarks) of 3,000 base pairs, and nearly 200 genes have been localized on the map. New genes can now be placed on this map in only a few hours. This process, which replaces mapping techniques that required weeks of effort, is now in use in more than 100 laboratories.

Despite this success, it has long been apparent that no straightforward extension of these methods would be successful on the greatly expanded scale of the human genome. For this reason, we developed a method to import manageable segments of the human chromosomes into yeast, where they can be propagated as yeast artificial chromosome (YAC) clones. One advantage of YACs over previous cloning systems is that there is no absolute upper size limit to the clones. We can now prepare large collections of YAC clones, each containing a different segment of human DNA, averaging hundreds of thousands of base pairs in size, a 10-fold improvement over the capacities of previous systems. Another advantage is that the methods of packaging, maintaining, and replicating DNA are more similar in yeast than in bacteria to the analogous processes in human cells. Consequently, a higher fraction of the human genome can be successfully propagated in yeast than in the more-conventional bacterial hosts.

Methods of recovering any desired segment of human DNA as a YAC have become standard during the past three years. YAC clones played a central role in two recent major successes in human genetics: discovery of the molecular basis of the fragile X syndrome, carried out in part in the laboratories of Thomas Caskey (HHMI, Baylor College of Medicine) and Stephen Warren (HHMI, Emory University); and the discovery of the gene that is mutated in familial adenomatous polyposis, carried out in part in the laboratory of Raymond White (HHMI, University of Utah). The fragile X syndrome is a common heritable cause of mental retardation; familial adenomatous polyposis is a genetic predisposition to colon cancer.

Now that YAC cloning is a proven method for recovering large blocks of human DNA, our attention has turned to the analysis of YACs. A need for efficiency arises from the sheer scale of human chromosomes: approximately 10,000 YACs would be required to recover the DNA present in all the human chromosomes. To meet this challenge, it will be necessary to develop a new area

of analytical biochemistry. Although the ultimate goal is fully automated methodology, the principles that underlie DNA analysis are too poorly understood to support immediate instrumentation development. One key will be to obtain a better grasp of the information flow that accompanies the mapping and sequencing of DNA. In collaboration with Will Gillett (Washington University), we have analyzed the computational problems posed by DNA-mapping methods in which large blocks of DNA are broken down into smaller segments by subcloning. Powerful, flexible software devoted to this problem is under development. Parallel efforts to achieve a better understanding of the experimental steps are also under way.

It is difficult to overestimate the extent to which the analysis of DNA—cloning, mapping, and sequencing—has come to underlie progress in biology and biomedical research. Biological research has been driven by advances in methodology, and the systematic analysis of DNA is one of the field's critical technologies. However, in proportion to the challenges posed by the genetic complexity and diversity of organisms, the development of this technology is still in its infancy.

Dr. Olson is now Professor of Molecular Biotechnology at the University of Washington, Seattle.

Molecular Genetic Studies of Hematopoietic Cells

Stuart H. Orkin, M.D.—Investigator

Dr. Orkin is also Leland Fikes Professor of Pediatric Medicine at Harvard Medical School. He received his B.S. degree in biology from the Massachusetts Institute of Technology and his M.D. degree from Harvard Medical School. His postdoctoral research was in the Laboratory of Molecular Genetics at NIH under the supervision of Philip Leder. Upon returning to Harvard, Dr. Orkin received specialty training in pediatric hematology at the Children's Hospital, where he later joined the faculty. His many honors include the Clinical Investigator Award from the American Federation for Clinical Research and the Dameshek prize of the American Society of Hematology. He is a member of the National Academy of Sciences.

ALL mature blood cells are derived from pluripotent hematopoietic stem cells, which constitute a rare population in the bone marrow. The decision of stem cells to differentiate leads to the production of a heterogeneous array of cells with varying developmental potentials and with commitment to expression of lineage-specific protein products. A major goal of this laboratory is an improved understanding of hematopoietic cell development and the expression and function of specific genes that relate to the normal biology of hematopoietic cells.

Efforts are directed to analyses of both red and white blood cells. These cell types are important in severe, clinically significant genetic disorders in which the capacity to produce specific proteins is impaired by mutation. In these studies we seek to describe the molecular basis of inherited disorders, understand the normal regulation of the affected genes, and utilize the findings from this basic work to formulate novel treatments based on molecular biologic considerations.

One of the major, classical disorders of red cells is β-thalassemia (also known as Cooley's anemia), in which the synthesis of hemoglobin is defective. Through molecular cloning and gene expression, the molecular basis of the disease was determined in this laboratory several years ago. Now the unsolved problems are related to how globin genes are normally regulated in developing erythroid precursor cells. Specifically, how are the globin genes activated only in red cells? How are different globin genes regulated in development? How do erythroid precursor cells arise during development from progenitor cells that have the potential to yield either red or white cells?

To approach these general problems, we have concentrated on identifying and characterizing unique DNA-binding proteins that appear to be major transcriptional regulators in erythroid cells. A prominent, apparently erythroid-specific DNA-binding protein was first discovered that recognizes a small DNA motif (GATA) found in the promoters or enhancers of all erythroid-expressed genes. Through molecular cloning, mammalian, avian, and amphibian homologues were characterized.

The protein is modular, consisting of a novel two-finger structure required for DNA-binding and other domains that serve as potent positive activators of gene transcription. Expression of the protein in two other hematopoietic cell types, megakaryocytes and mast cells, suggests that it is first expressed in a multipotential progenitor cell and may regulate genes in those cell types as well. Recent data have supported these conclusions. Attention has been directed to how this transcription factor is itself regulated in hematopoietic cells. An improved understanding may provide important insights into the initial events involved in erythroid decision-making and maturation.

Studies of the gene revealed an element in the promoter region that serves as a site for positive autoregulation. In this manner, expression of the factor tends to maintain its own expression and stabilize the differentiated state. Furthermore, the promoter for the receptor for the erythroid-specific growth factor, erythropoietin, is under control by this transcription factor. By such circuitry, expression of the factor tends to guarantee subsequent erythroid development and viability.

Site-specific gene disruption in mouse embryo-derived stem cells and generation of chimeric animals has also revealed that the protein is essential for normal erythroid development and that related proteins that bind the GATA motif cannot compensate for its absence. Using *in vitro* differentiation of embryo-derived stem cells into hematopoietic cell types, we have developed an experimental system that permits assessment of the role of GATA-transcription factor in erythroid development and systematic testing of various aspects of the function and/or regulation of the protein.

In separate but conceptually related studies, a gene that encodes an essential component of the white blood cell (phagocytic) system responsible for killing ingested microorganisms is being examined in an effort to understand how this

clinically important host defense system is regulated and, more generally, how cell-specific gene expression is achieved in this lineage, also descendent from the pluripotent stem cell. The gene under study encodes a subunit of a unique cytochrome that is defective in an X-linked condition, chronic granulomatous disease. By positional cloning, we previously isolated the gene, determined its structure, and demonstrated the presence of the protein product in the cytochrome complex of phagocytic cells.

In addition, since interferon-γ stimulates phagocytic cell function generally as well as expression of the cytochrome, it was also possible to show that this lymphokine is clinically effective in chronic granulomatous disease. Studies have identified several point mutations in the cytochrome that interfere with protein function *in vivo* and mutations in an associated cytochrome subunit encoded by an autosome in rarer cases of chronic granulomatous disease.

More current efforts are directed toward defining the elements of the gene responsible for approaching regulation in phagocytic cells. Through the use of transgenic mice, we identified a DNA fragment sufficient for targeting reporters and oncogenes to a subset of phagocytes *in vivo*. Coupled to an oncogene, this fragment leads to the consistent development of an inherited malignancy of phagocytes in mice.

In the promoter we identified a negative regulatory site that binds a protein whose concentration falls dramatically during phagocyte differentiation. This factor corresponds to a putative repressor termed CCAAT displacement protein (CDP). Through molecular cloning, we have characterized human CDP and shown that it is highly related to a novel *Drosophila* homeobox protein, a product of the *cut* gene, which controls cell fate decisions in several tissues during fly development. Although this repressor does not explain the restriction of cytochrome expression to phagocytic cells, its loss combined with the presence of unknown positive, white cell–specific regulators very likely accounts for the temporal pattern of cytochrome expression during cellular maturation.

More generally, the remarkable similarity of human CDP and the fly *cut* protein suggests functional correlates and predicts that CDP will participate in cell fate decisions during mammalian embryogenesis. Current efforts are directed to analysis of the role and function of CDP in mammalian cells and to identification of the critical cis- and trans-components responsible for white cell–specific gene expression.

Albinism and Tyrosinase

Paul A. Overbeek, Ph.D.—Assistant Investigator

Dr. Overbeek is also Associate Professor of Cell Biology at Baylor College of Medicine and has adjunct appointments in the Institute for Molecular Genetics, the Division of Neuroscience, and the Department of Ophthalmology at Baylor College of Medicine. He received his B.A. degree in chemistry from Kalamazoo College, his Ph.D. degree in cellular and molecular biology from the University of Michigan, and an M.B.A. degree from the University of Chicago. His postdoctoral research was done in the laboratory of Heiner Westphal at NIH.

THE genetic disorder of albinism, or loss of pigmentation, has been identified in many species. Albinism typically is caused by the loss of production of the black pigment termed melanin. Specialized cells known as melanocytes are responsible for melanin synthesis in the skin, hair, and iris. In addition to melanocytes, there are cells at the posterior of the retina, referred to as retinal pigment epithelial cells, that normally synthesize melanin. When these two types of cells lose their melanin-synthesizing ability, albinism results. Since melanin helps protect the skin from ultraviolet radiation and the visual system from bright light, albinism is often associated with secondary health problems, including increased risk of skin cancer and visual system deterioration.

Melanin is derived from the amino acid tyrosine. An enzyme known as tyrosinase converts tyrosine to dopaquinone, which is then polymerized to produce melanin. Studies of albino individuals, including humans and laboratory mice, have revealed that albinism in often associated with a decrease or loss of tyrosinase enzymatic activity. This observation has led to the hypothesis that albinism may be caused by a mutation in the gene that encodes tyrosinase. A mutation could result in synthesis of an abnormal tyrosinase protein that no longer has enzymatic activity. We have recently undertaken a series of experiments to test this hypothesis in albino laboratory mice.

One important prediction of this hypothesis is that the nucleic acid sequence of the tyrosinase gene in pigmented mice will be different from the sequence in albino mice. A procedure known as the polymerase chain reaction was used to amplify specific regions of the tyrosinase gene from more than 20 different pigmented and albino mouse strains. The nucleic acid sequences were then determined, and a computer program was used to compare them. Interestingly, every pigmented strain was found to encode a cysteine at amino acid 103 of the tyrosinase gene, while the albino mice all had a single base pair change that would cause amino acid 103 to become a serine. These results suggested that conversion of cysteine to serine at amino acid 103 was sufficient to inactivate the tyrosinase enzyme and produce albinism.

In order to confirm this prediction, we used recombinant DNA techniques to construct two miniature versions of the tyrosinase gene: one with a cysteine, the other with a serine at amino acid 103. The two constructs were tested in an albino strain of mice. Embryos of the albino strain were harvested at the one-cell stage and injected under the microscope with purified DNA. After injection the embryos were reimplanted into the reproductive tracts of pregnant females and allowed to develop to birth.

The newborn mice were then screened to identify those mice (termed transgenic) in which the injected DNA had become stably incorporated into the genome. Mice that had incorporated the cysteine 103 version of tyrosinase became pigmented, while all of the mice that had integrated the serine 103 version of tyrosinase were still albino. These experiments have confirmed that albinism can be caused by a mutation in the tyrosinase gene and that a single base pair substitution is sufficient to inactivate the gene.

These studies of albinism have provided an important new strategy for the identification of transgenic mice. Prior to the design of the miniature tyrosinase gene, transgenic mice were identified by techniques that required genomic DNA isolation. The miniature tyrosinase gene allows transgenic mice to be recognized by simple visual inspection for pigmentation.

To test the general usefulness of this system, experiments were done in which two different recombinant DNAs were co-injected into embryos. One was the miniature tyrosinase gene; the other was designed to generate a mouse model for predisposition to kidney cancer. The newborn mice were first checked for pigmentation, then genomic DNA was isolated and tested for integration of the co-injected DNA. Altogether five pigmented mice were obtained, and all five had integrated both DNAs. Breeding studies showed that the two different DNAs were always transmitted together.

When the kidneys were checked for cancer in-

duction, only the pigmented mice were found to have kidney lesions. Neoplastic changes were limited to the kidney and did not occur in the skin. These co-injection experiments demonstrate that two unrelated genes can become located in close proximity to each other in the genome and still function in an independent fashion. Moreover, this strategy simplifies long-term maintenance of experimental transgenic mice, since the desired mice can be readily identified by their pigmentation.

An additional question was whether the synthesis of tyrosinase in some cell type other than melanocytes or retinal pigment epithelial cells would lead to melanin synthesis and pigment formation. To answer this question, regulatory sequences that were known to be active specifically in lens cells were linked to the tyrosinase coding sequences using recombinant DNA techniques, and transgenic mice were generated in an albino strain of mice. The transgenic mice all developed black eyes. Synthesis of tyrosinase in the lens led to pigment production in the lens, confirming that the reason lens cells do not normally make melanin is because they do not normally make the enzyme tyrosinase. Previous studies had suggested that pigment synthesis in inappropriate cell types might be harmful for those cells. In the transgenic mice, the pigmented lens cells show histological evidence of cellular injury and inhibition of normal growth.

When transgenic DNA integrates into the genome, it can insert into the middle of an endogenous gene, causing insertional inactivation of the gene. Since our transgenic tyrosinase mice were easy to identify and breed, 80 families were tested for the presence of insertional mutations. Eight families were found that have defects, ranging from embryonic lethality to male sterility to anemia and premature kidney failure.

One of the mutations is particularly fascinating because the homozygous transgenic mice all have an inversion of the left-to-right organization of their internal organs. This condition in humans is known as *situs inversus*. Since the factors that control polarity in the developing mammalian embryo have not yet been identified, we have begun experiments that will make use of the transgenic insert to try to isolate the regulatory factor for *situs inversus*. The identification and characterization of such a factor could greatly enhance our understanding of embryonic development and guide future efforts to help prevent analogous birth defects in humans.

Structural Studies of DNA-binding Proteins

Carl O. Pabo, Ph.D.—Investigator

Dr. Pabo is also Professor of Biophysics and Structural Biology at the Massachusetts Institute of Technology. He received his undergraduate degree from Yale University, where he majored in molecular biophysics and biochemistry. He did his graduate work in Mark Ptashne's laboratory at Harvard University, where he continued his research as a Jane Coffin Childs fellow in the laboratories of Stephen Harrison and Don Wiley. Prior to accepting a position at MIT, Dr. Pabo was Professor of Molecular Biology and Genetics and of Biophysics at the Johns Hopkins University School of Medicine.

We are interested in understanding how proteins recognize specific sites on double-stranded DNA and how the bound proteins regulate gene expression. We would like to know what structural motifs are used by DNA-binding proteins, what side chains make sequence-specific contacts, and whether there are any recurring patterns or rules for recognition of sites on double-stranded DNA. Much of our current research has focused on characterizing the major structural motifs found in DNA-binding proteins. We hope to use this information to design novel DNA-binding proteins for research, diagnosis, and therapy.

Prokaryotic repressors provide useful model systems for the study of protein-DNA interactions, and we are continuing to study several bacterial repressors. The repressor from the bacteriophage λ uses a helix-turn-helix motif and an extended amino-terminal arm to contact sites in the major groove. The *arc* repressor from *Salmonella* bacteriophage P22 uses a β-sheet for site-specific recognition. The major developments in our laboratory during the past two years, however, have involved studies of two of the key motifs—the homeodomain and the zinc finger—that are used by eukaryotic regulatory proteins.

Crystal Structures of Homeodomain-DNA Complexes

The homeodomain is a conserved structural motif found in many eukaryotic proteins that regulate development and cell fate. To understand how this motif recognizes DNA and how this is related to the helix-turn-helix motif seen in prokaryotic repressors, we have determined the crystal structures of two homeodomain-DNA complexes.

We began by studying the homeodomain from the engrailed protein, which plays a key role in *Drosophila* development. (This project is a collaboration with Thomas Kornberg at the University of California, San Francisco.) We were able to grow good cocrystals of the homeodomain-DNA complex, and Chuck Kissinger solved the structure of this complex. The homeodomain makes contacts in both the major and minor grooves. The helix-turn-helix unit makes critical contacts in the major groove, but the orientation of this helix-turn-helix unit with respect to the DNA is different than the arrangements observed with the prokaryotic repressors. Residues near the amino-terminal end of the homeodomain form an extended "arm" that fits into the minor groove and makes additional site-specific contacts.

We also have been studying a complex containing the homeodomain from the α2 protein, which helps to regulate mating type in yeast. (This project is a collaboration with Alexander Johnson at the University of California, San Francisco.) Cynthia Wolberger recently solved this structure, and comparison with engrailed revealed that 1) the structures of these two homeodomains are very similar (despite a 3-residue insertion in α2 and despite significant amino acid sequence differences) and 2) the orientation of the helix-turn-helix unit with respect to the DNA also is conserved. This conserved docking arrangement is maintained by side chains that are identical in α2 and engrailed. Because these residues tend to be conserved among all homeodomains, these structures may provide a general model for homeodomain-DNA interactions.

Our studies of the homeodomain are supported by a grant from the National Institutes of Health.

Crystal Structures of Zinc Finger–DNA Complexes

The zinc finger domain, which contains about 30 amino acids, is another key DNA-binding motif that is found in a large family of eukaryotic regulatory proteins. Studies from other groups have shown that each finger contains an antiparallel β-sheet and an α-helix, but little has been known about how these fingers recognize DNA. Nikola Pavletich recently solved the structure of a complex containing three zinc fingers from a murine transcription factor. Starting with cDNA for the *zif268* gene (provided by Daniel Nathans, HHMI, Johns Hopkins University), he cloned and expressed a three-finger peptide and crystallized the peptide-DNA complex. The zinc fingers rec-

ognize B-DNA and fit into the major groove. Each finger makes its primary contacts with a 3–base pair "subsite," and side chains near the amino-terminal end of the α-helix make the critical contacts with the bases.

Since the zinc fingers are used in a modular fashion, they may be the ideal motif to use as we try designing novel DNA-binding proteins. However, we need to determine the structure of additional zinc finger–DNA complexes so that we can see how this motif is used in other proteins. (Do all zinc fingers dock against the DNA in the same way?) Nikola Pavletich has cloned the zinc finger regions from a number of other proteins and has recently solved the structure of a complex that contains the five zinc fingers from GLI, a protein that is amplified in a subset of human tumors. Finishing this structure and comparing it with the zif complex should provide a firmer basis for the design projects.

Structural Studies of Other Protein-DNA Complexes

It is important to obtain structural information about the other major DNA-binding motifs that occur in eukaryotic regulatory proteins. We are focusing on the helix-loop-helix proteins and the POU domain, which have very important roles in development. We also are working with the TFIID protein (because of its central role in transcription) and with the p53 protein, which is the most common site of mutations in human tumors.

Because structural analysis often is limited by the ability to obtain suitable crystals, we are trying to improve methods for the cocrystallization of protein-DNA complexes. Our initial approach involved systematic changes in the length of the DNA site and required that the entire site be resynthesized for each experiment. We now have encouraging preliminary results with a linker cocrystallization scheme that combines the protein, the binding site, and a library of DNA linkers that can be used with any complex. This strategy may allow a dramatic increase in the number of cocrystallization conditions that can be tested.

Design of Novel DNA-binding Proteins

We are attempting to use the zinc finger structures as a basis for designing novel DNA-binding proteins. Two major approaches are being tested: 1) genetic strategies for selecting zinc finger proteins that recognize desired target sequences and 2) strategies for computer-aided protein design. These programs, which can systematically consider a large number of sequences and conformations, are being used in our attempts to design zinc finger proteins that will recognize novel binding sites. This combination of structural analysis, computer-aided protein design, and genetic selection should provide a better understanding of protein-DNA recognition and allow rapid design of zinc finger proteins that recognize novel target sites.

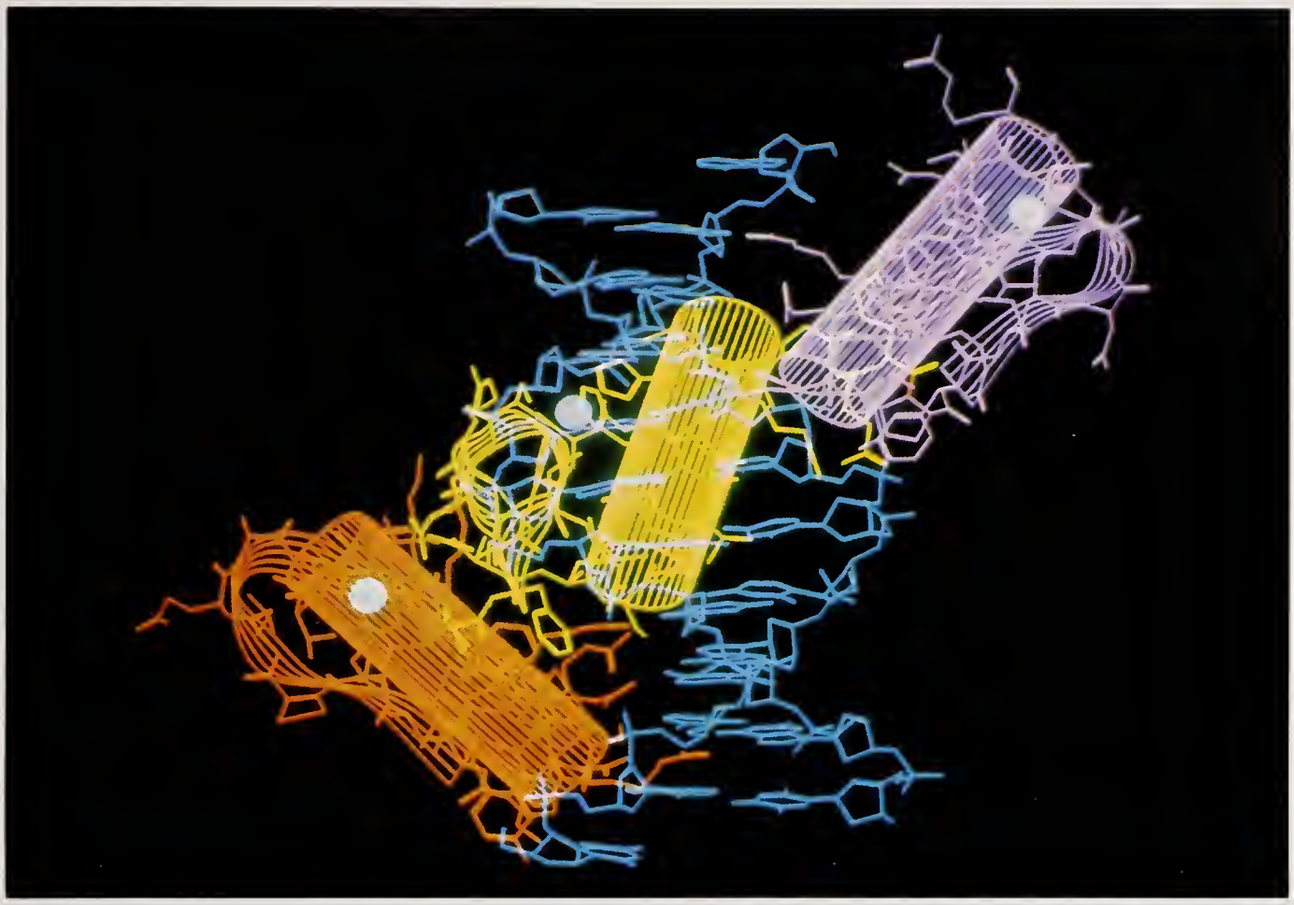
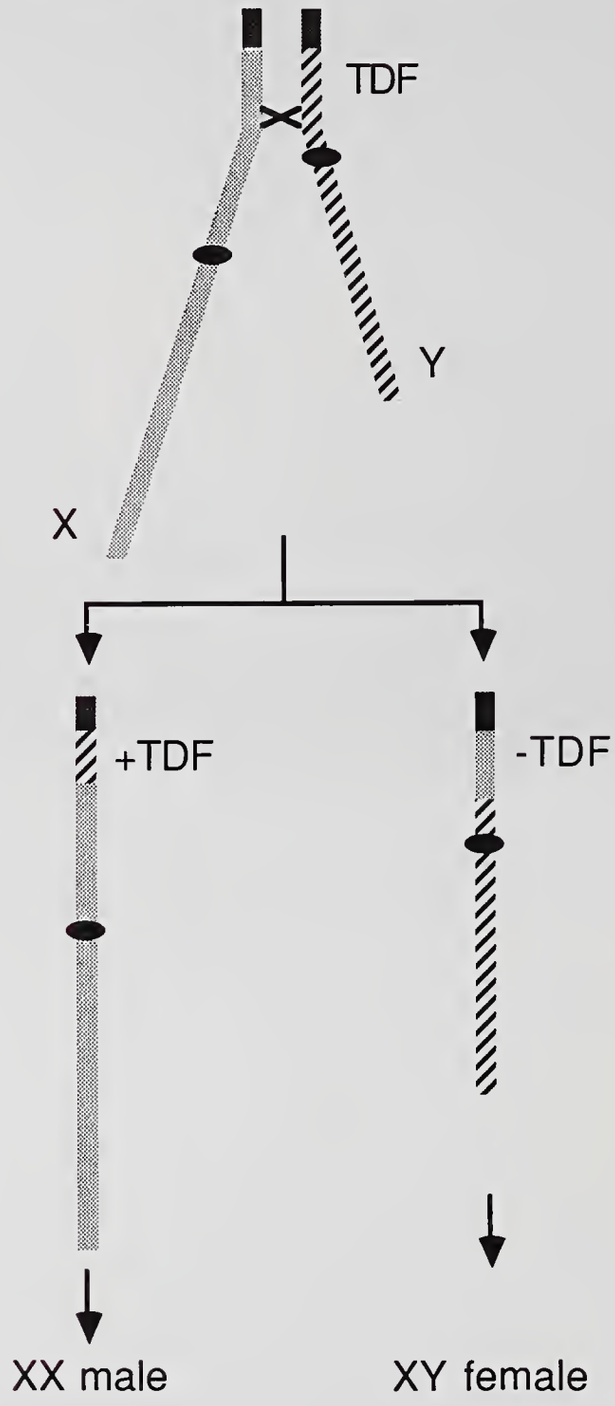


Image of an x-ray crystal structure showing how regulatory proteins of a key class known as zinc finger bind to DNA. The protein contains three "fingers" (shown in orange, yellow, and purple) that bind to the double-helical DNA (blue). Each finger has an α-helix that fits into the major groove of the DNA and makes critical contacts that help the protein recognize its proper binding site.

From Pavletich, N.P., and Pabo, C.O. 1991. Science *252:809–817. Copyright © 1991 by the AAAS.*

Hypothesis: XX Males and XY Females Result from X-Y Interchange at Paternal Meiosis



The X-Y interchange hypothesis: XX males and XY females receive reciprocal products of similar aberrant X-Y exchanges occurring in the fathers during the production of sperm (meiosis). TDF is testis-determining factor, the sex-determining gene normally located on the Y chromosome.

Research of David Page.

The X and Y Chromosomes in Mammalian Development

David C. Page, M.D.—Assistant Investigator

Dr. Page is also Associate Member of the Whitehead Institute for Biomedical Research, Associate Professor of Biology at the Massachusetts Institute of Technology, and Assistant Professor at the Harvard University–MIT Division of Health Sciences and Technology. He received his undergraduate degree in chemistry from Swarthmore College and a medical degree from Harvard Medical School and the Harvard-MIT Health Sciences and Technology Program. After training with Raymond White, at the University of Massachusetts, and David Botstein, at MIT, Dr. Page became one of the first Fellows of the Whitehead Institute. He subsequently joined the faculties of Whitehead and MIT.

TO a large degree, human individuals differ in physical characteristics because of the impact of genetic variation on the course of embryonic development. The human genome is organized into 23 pairs of chromosomes, each believed to carry, on average, about 5,000 genes. When considering genetic differences among individuals, it is important to distinguish between variation in a single gene, called Mendelian, and massive multigene variation, as found in chromosomal disorders.

As discussed elsewhere in this volume, single-gene defects are responsible for certain conditions such as color blindness, cystic fibrosis, and muscular dystrophy. Other conditions, such as Down syndrome, appear to be the result of "wholesale" abnormalities affecting an entire chromosome. As reductionists and molecular biologists, we proceed on the assumption that the developmental consequences of chromosomal abnormalities will ultimately be understood in terms of individual genes and their particular functions.

My colleagues and I are seeking to understand how massive variability in one chromosome pair —the sex chromosomes—dramatically affects the course of development. Embryos normally inherit one sex chromosome from each parent. The mother contributes an X chromosome, and the father contributes either an X or a Y. Thus normal embryos have one of two sex chromosome constitutions, XX or XY.

Sex Determination

In both humans and mice the presence or absence of the Y chromosome determines whether an embryo develops as a male or as a female. XX embryos become females, XY embryos males. For years scientists wondered whether the Y chromosome carried few or many sex-determining genes and how those were distributed along the chromosome.

We found that the entire sex-determining function can be traced to one tiny portion of the human Y chromosome. This sex-determining region was identified by studying DNA from "XX males" and "XY females." XX males have small testes and are sterile. XY females are also sterile and do not develop secondary sexual characteristics. We found that almost all XX males had inherited a small bit of the Y chromosome attached to one of their X chromosomes. Conversely, some XY females lacked the same segment of the Y that was present in XX males. On the basis of the chromosomal deletions found in such patients, we constructed a map of the Y chromosome. It was then we came to recognize that the presence or absence of one small region, about 0.4 percent of the chromosome, correlated well with gender.

Detailed analysis of XX males suggests that one or more genes within this relatively small segment of the Y chromosome determine the outcome of sexual development. Laboratories around the world have scoured this relatively small segment of the Y chromosome searching for such sex-determining genes. We are now conducting a variety of experiments to characterize the functions of two genes in the region: *ZFY*, which we identified in 1987, and *SRY*, a gene described by British scientists in 1990.

Both *ZFY* and *SRY* appear to encode DNA-binding proteins that are likely to regulate the transcriptional activity of particular but unknown target genes. Experiments clearly demonstrate that *SRY* is a sex-determining gene. Much less clear is the role, if any, that *ZFY* plays in the process. We hope to understand better the function of the *ZFY* gene by simultaneously analyzing a closely related gene, *ZFX*, that we identified on the X chromosome.

In some human XY females, we have identified mutations near or within the *SRY* gene. By contrast, other XY females, some human and some mouse, appear to have intact Y chromosomes but may have mutations elsewhere, perhaps in autosomal genes that play important roles in sex determination. Identification of such autosomal sex-determining genes is a future goal. Our studies of sex determination are supported by a grant from the National Institutes of Health.

Turner Syndrome

As mentioned earlier, embryos normally have two sex chromosomes. However, about 1–2 percent of all human embryos have only one. The vast majority of such XO embryos are lost to spontaneous miscarriage, but a few survive. The surviving XO embryos develop as females with a particular set of physical features known as Turner syndrome, which includes short stature, webbing of the neck, puffiness of the hands and feet, and failure of secondary sexual development. It had been postulated that Turner syndrome might be the result of having a single copy of one or more genes common to the X and Y chromosomes. Nothing was known as to the number or nature of these hypothetical Turner genes.

We began to focus our attention on this disorder when it was noticed that certain XY females exhibit the same anatomic abnormalities as XO females. A pivotal finding was that all such XY Turner females lacked a portion of the Y chromosome. We postulated that the Y chromosomal deletions in these individuals might encompass not only a sex-determining gene or genes but also a nearby Turner gene or genes.

Pursuing this hunch, we discovered two candidate Turner genes, one on the Y chromosome and one on the X chromosome. These genes, named *RPS4Y* and *RPS4X*, appear to encode slightly different forms of a protein constituent of the ribosome, a structure required for protein synthesis and vital to all cells. In embryos lacking a second *RPS4* gene (i.e., having a single *RPS4*), the rate at which ribosomes are constructed may be slowed, in turn reducing the embryo's capacity to synthesize other proteins. We are currently testing the highly speculative hypothesis that such a reduction in protein synthetic capacity is the cause of some of the physical features of Turner syndrome.

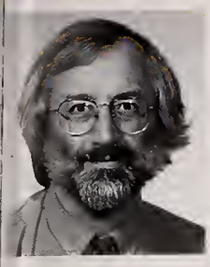
An interesting analogy can be found in the fruit fly *Drosophila melanogaster*. There, deficiencies in ribosomal protein genes are associated with a particular "syndrome" called the "Minute" (pronounced mi-NUTE) phenotype, which includes reduced body size, diminished viability and fertility, and specific anatomic abnormalities. The *Drosophila* Minutes may serve as a useful model system in which to study effects of *RPS4* gene dosage. We are also exploring potential mouse models of Turner syndrome.

The very existence of related but nonidentical ribosomal protein genes on the X and Y chromosomes raises the possibility that the ribosomes of human males may differ slightly from those of females. Experiments still in their early stages suggest that this is the case. Thus the differences between the sexes may extend all the way down to the most fundamental and vital of intracellular machines!

The Human Y Chromosome

As mentioned above, we constructed a map of the human Y chromosome by characterizing naturally occurring deletions, such as those found in XY females and XX males. We are continuing to refine this map, which is useful not only in studying sex determination and Turner syndrome, but also in examining the role of Y chromosomal genes in other processes, including the development of certain cancers and the making of sperm. We recently set out to "clone" the human Y chromosome as a series of overlapping segments, each segment constituting about 1 percent of the chromosome. Such an ordered array of cloned segments should facilitate the process of locating and characterizing all genes on the Y chromosome. Our efforts to map and clone the Y chromosome are supported by a grant from the National Institutes of Health.

Mammalian Development and Disease



Richard D. Palmiter, Ph.D.—Investigator

Dr. Palmiter is also Professor of Biochemistry at the University of Washington. He received his Ph.D. degree from Stanford University and did postdoctoral work at Stanford, Searle Research Laboratories in England, and Harvard University. Prior to his current work with transgenic animals, Dr. Palmiter studied the mechanism of steroid hormone action in the chick oviduct and the regulation of metallothionein gene expression in mice. He is a member of the National Academy of Sciences and the American Academy of Arts and Sciences.

ABOUT 10 years ago we began a fruitful collaboration with Ralph Brinster's laboratory at the University of Pennsylvania. Together we helped develop methods for introducing functional genes into all cells of the mouse. The genes under study are manipulated and grown in bacterial plasmids, using standard recombinant DNA techniques. Then the regions of interest are excised from the plasmid, and a few hundred copies are injected into the pronucleus of a fertilized mouse egg (or that of any other mammal).

Remarkably, the DNA integrates about 30 percent of the time into one of the chromosomes prior to replication, and the genes are inherited by all daughter cells, as any other gene would be. Furthermore, many of the genes are functional, imparting new genetic characteristics to the animal. Mice and other animals carrying foreign DNA are referred to as transgenic. Because the new genes are also in the germ cells, they are usually transmitted to subsequent generations.

One of our goals has been to discover what parts of a gene determine when, where, and how efficiently it will be utilized. We often start by testing a large piece of DNA that includes the gene of interest. In transgenic animals, the gene will usually be expressed in the appropriate time and place, even though it has integrated at an abnormal chromosomal location and may be derived from a different mammalian species. Then we delete various regions of the genes, and with each variant we make transgenic mice to determine what regions are essential for appropriate expression.

For example, we have delineated a small region (125 base pairs) of the rat elastase I gene that is essential for the expression of the gene in acinar cells of the pancreas. Furthermore, this sequence (often called an enhancer) can be used to direct the expression of another gene (e.g., the growth hormone gene) to the acinar cells, and the sequence will function when positioned almost anywhere in the vicinity of the growth hormone gene. In similar experiments we have been identifying sequences responsible for directing appropriate expression of globin genes in red blood cells, albumin in hepatocytes, and protamine I in male germ cells.

Because regulatory elements from one gene can often be used to control another; the expression of many interesting genes can be directed to a specific cell type and the consequences on the development and function of those cells can be assessed. For example, using the elastase enhancer element, we have been able to make strains of mice that reproducibly develop pancreatic cancer as a consequence of expressing the transforming gene from simian virus 40, the mouse *myc* gene, or the human H-*ras* oncogene. Similarly, we have developed models of liver cancer by directing the expression of these genes to hepatocytes with the albumin enhancer.

Significantly, each of these genes results in a characteristic morphological transformation of the organ, which probably reflects the particular cellular events that these genes mediate. By means of simple genetic crosses, mice carrying any pair of these transforming genes can be created. They develop tumors that appear more rapidly and grow more aggressively than those in mice carrying a single gene, suggesting that these genes act cooperatively.

Some genes that are not generally considered oncogenes may also predispose cells to malignant transformation and cancerous growth. In one example, we expressed the surface antigens of hepatitis B virus (HBV) in the liver, using the albumin enhancer. In transgenic mice, expression of this gene resulted in the synthesis of the viral surface antigen and envelope protein, which aggregated within the secretory apparatus of the liver cells, causing cellular injury and death. This is accompanied by liver cell regeneration. However, when the mice were more than a year old, they developed liver cancer. Because HBV infects millions of people worldwide, and the incidence of liver cancer among them is very high, this result may indicate that chronic expression of HBV surface antigens may be a contributing factor. Similarly, expression of plasminogen activator in the liver using the albumin enhancer results in defective blood clotting and liver injury. In this case,

liver cells that have deleted the transgene repopulate the liver, and when the mice are a few months old they seem normal. However, they also succumb to liver cancer when they are one to two years old. In both cases we postulate that liver regeneration in the toxic environment of liver injury results in genetic damage that predisposes the liver cells to malignant transformation.

It is also possible to develop transgenic mice that mimic some human genetic diseases. For example, we made a model of human sickle cell disease. By introducing into mice both human α- and β-globin genes under control of the locus control region (a newly discovered genetic element essential for high-level expression of globin genes), we have made mice that produce as much human hemoglobin as mouse hemoglobin. When the mutant β-globin gene from people with sickle disease is substituted for the normal gene in these experiments, the red blood cells of the mice sickle under appropriate conditions. These mice may be a valuable resource for testing experimental therapies.

A long-range goal is to use transgenic mice to study aspects of neural development. We have started by cloning the genes involved in the synthesis of the catecholamine neurotransmitters: dopamine, norepinephrine, and epinephrine. The control elements from these genes are being tested in conjunction with genes whose products can be easily visualized, to assess when and where they are expressed during development. For example, we have shown that the regulatory elements from the gene responsible for making norepinephrine direct the expression of β-galactosidase to certain neurons of the central nervous system, the peripheral nervous system, the enteric nervous system, and the adrenal medulla (see figure). Because this marker gene is expressed very early during neural development, it allows us to visualize the cells while they are still migrating to their final destinations and before they acquire properties of mature neurons.

By mating these mice to mice carrying a genetic defect that affects the innervation of the bowel and results in a condition similar to Hirschsprung's disease in humans, we have shown that neuronal precursors fail to migrate into the distal portion of the gut. The lack of innervation of the colon results in chronic impaction of fecal material and ultimately death.

We are also using the control elements from these genes to direct the expression of other genes (e.g., encoding neurotransmitters, growth factors, hormones, proteases, or oncogenes) to these neurons with the aim of affecting the decisions they make during the process of forming functional connections with target cells.

Blue staining reveals the location of cells that express a β-galactosidase transgene in neural precursors in a 10.5-day mouse embryo.

Research and photograph by Raj Kapur in the laboratory of Richard Palmiter.

Regulation of Gene Expression in Steroid Hormone Biosynthesis

Keith L. Parker, M.D., Ph.D.—Assistant Investigator

Dr. Parker is also Associate Professor of Medicine and Biochemistry at Duke University Medical Center. After attending Williams College, he earned his M.D. and Ph.D. degrees in genetics at Washington University, studying with Donald Shreffler. He served as intern and resident in internal medicine at Parkland Memorial Hospital, Dallas. He then moved to the Department of Genetics at Harvard Medical School, where he was a postdoctoral fellow with Jonathan Seidman. Dr. Parker's next move was to the faculty of Duke University Medical Center.

THE adrenal gland plays an essential role in the body's ability to respond to stress. Two different parts of the adrenal gland, an outer cortex and an inner medulla, produce distinct components of this response. The medulla produces epinephrine and norepinephrine, which are released very rapidly, preparing the organism for immediate physical activity. In contrast, the cortex produces steroid hormones, which are released more slowly and exert prolonged effects.

These adrenal steroids constitute two major classes: glucocorticoids, which are made in the inner zone of the cortex and control carbohydrate metabolism, and mineralocorticoids, which are made in the outer zone and regulate salt and water balance. Both classes of steroid hormones are formed from cholesterol by the sequential action of a related group of steroidogenic enzymes. One of these, the cholesterol side-chain cleavage enzyme, is expressed in all steroidogenic tissues. A second, 21-hydroxylase, is expressed throughout the adrenal cortex. Finally, there are distinct forms of 11β-hydroxylase: one form produces mineralocorticoids in the outer zone, and the other produces glucocorticoids in the inner zone. The physiological regulators of these two classes of adrenal steroids differ markedly, despite the shared role of the same enzymes in their biosynthesis.

We are interested in defining the events that control the expression of the adrenal steroidogenic enzymes. These studies have addressed two major questions. First, what mechanisms direct the expression of these related genes within adrenocortical cells? Second, what determines the functional differentiation of the adrenal cortex into mineralocorticoid- and glucocorticoid-producing zones?

Our studies of gene regulation have focused on the 5′-flanking regions of these genes. This part, the promoter region, contains most sequences important in transcriptional regulation of other genes. These promoter analyses identified a protein, steroidogenic factor 1, that plays a major role in regulating the expression of all three steroidogenic enzymes. We only found this protein in cells that made steroid hormones, suggesting that it contributes to the cell-selective expression of these genes. The global role of this protein in the expression of three distinct genes suggests that it coordinates the adrenocortical expression of a network of enzymes.

We next used cow adrenal glands, providing much greater amounts of protein, and purified the bovine form of steroidogenic factor 1. The presently available amounts of purified protein should be sufficient to determine its amino acid sequence and to raise specific antibodies. The combination of specific antibodies and amino acid sequence data should allow us to clone the gene encoding this key regulatory protein. By comparing the primary structures of steroidogenic regulatory protein and the previously described transcriptional regulatory proteins (such as the steroid hormone receptor proteins), we may gain new insights into the mechanisms that regulate the adrenal steroidogenic enzymes. The specific probes and antibodies to steroidogenic factor 1 will further permit us to study the mechanisms that regulate its expression. These studies will provide new insights into the basis for tissue-specific differences in the production of steroid hormones and may identify important avenues for therapeutic intervention in clinical settings of abnormal adrenal steroidogenesis.

In a related effort, we are trying to define the potential role of the steroidogenic enzymes in hypertension. Extremely prevalent, hypertension affects approximately 20 percent of the adult population and is a major risk factor for heart attacks and strokes, the leading causes of death in developed nations. Although the underlying defect is unknown in most cases, family studies indicate a significant genetic component. Certain animal models of hypertension have directly implicated disordered adrenal steroidogenesis as an important contributor to hypertension. We are therefore investigating in more detail the role of adrenal steroids in hypertension.

Initial studies implicated a single 11β-hydroxylase protein in the biosynthesis of both mineralocorticoids and glucocorticoids, but we

subsequently defined two mouse 11β-hydroxylase genes located on chromosome 15. Although the overall structures of the two genes are quite similar, selected regions diverge significantly, suggesting that the proteins encoded by these two genes might have different activities. Moreover, the 5′-flanking regions of the two genes diverge significantly, suggesting that their regulation may differ.

To assess the enzymatic activities of the proteins produced by these two genes, we analyzed their activities following gene transfer into COS-7 cells, which normally do not make any steroid-metabolizing enzymes. One of the two proteins, designated 11β-hydroxylase, was able to produce glucocorticoids but could not make mineralocorticoids. In contrast, the other protein, designated aldosterone synthase, performed all of the reactions needed to make aldosterone. Thus there are significant differences in the enzymatic activities of the two mouse 11β-hydroxylase homologues.

To see where these two proteins are made in the adrenal cortex, we used the technique of *in situ* hybridization to examine sites of expression in the sections of the mouse adrenal gland. The aldosterone synthase protein is only present in the outer zone of the cortex, where mineralocorticoids are made. In contrast, the 11β-hydroxylase protein is only present in the sites of glucocorticoid production. These results document an exquisite coupling between enzymatic activities and sites of expression. The ability to maintain this segregation of the two proteins is undoubtedly an important part of the adrenal gland's ability to regulate separately the production of mineralocorticoids and glucocorticoids.

Based on these results, we plan to use these promoter regions in transgenic mice to target gene expression to specific cortical zones. Initially we are using the two promoters to direct the expression of renin, a gene previously shown to cause genetic hypertension when expressed in multiple tissues of transgenic animals. If successful, these experiments will validate the zone-specific expression of the two promoter regions and will establish that adrenal expression of renin is relevant to this hypertensive state.

We will also express the aldosterone synthase protein in the inner, glucocorticoid-producing zone. To this end, we have prepared a hybrid gene with the inner-zone–specific promoter driving expression of the aldosterone synthase gene. Based on the gene transfer experiments described above, we anticipate that this hybrid gene will synthesize large amounts of mineralocorticoids, thereby creating a genetic form of hypertension. Moreover, treatment with glucocorticoids should alleviate the hypertension by suppressing the expression of the hybrid gene in the inner zone. Recent studies strongly suggest that just such a mechanism is responsible for a subset of human patients with a glucocorticoid-remediable hypertension.

Molecular Neuroimmunology

Donald G. Payan, M.D.—Assistant Investigator

Dr. Payan is also Associate Professor of Medicine and of Microbiology and Immunology at the University of California, San Francisco. He received his B.S. degree in physics and mathematics from Stanford University. He went on to do graduate work in physics at the Massachusetts Institute of Technology and then returned to Stanford Medical School, where he received his M.D. degree. His medical residency at Massachusetts General Hospital, Boston, was followed by fellowships in infectious diseases at MGH and in allergy-immunology at Brigham and Women's Hospital.

MY laboratory continues to study the interactions between the nervous and immune systems at the molecular level. Ongoing work is focusing on the signal transduction pathways that are activated in cells into which we have transfected a number of neuropeptide receptors, in particular the tachykinin receptor for substance P. An additional effort is directed at understanding the biochemical properties of agrin, the nicotinic acetylcholine receptor–clustering molecule, which was cloned and sequenced during sabbatical studies with Richard Scheller (HHMI, Stanford University).

The two main projects currently under way are the analysis of the signal transduction capabilities of the substance P receptor (SPR) and accessory molecules that mediate that signaling, and an analysis of agrin's protease inhibitor domains and their potential function in modulating neuronal plasticity in the developing brain. With Julie Sudduth-Klinger, Christine Christian, and Mark Gilbert, we have transfected SPR cDNA into a number of different immune cells, in particular Jurkat and Reh, in order to study the effects of this receptor's expression on the cells' immunologic properties. We can get approximately 100,000 functional receptors per cell and can demonstrate a very brisk mobilization of intracellular calcium and inositol phosphate (IP_3) metabolites following stimulation with substance P.

Of great interest are many differences we have observed in comparing the Jurkat SPR with the Jurkat muscarinic receptor. Activation of the muscarinic receptor results in desensitization of subsequent T cell receptor activation with an anti–T cell receptor antibody; conversely, activation of the T cell receptor does not result in desensitization of the transfected muscarinic receptor. In contrast, we find that activation of Jurkat SPRs by substance P does not desensitize subsequent T cell activation, but activation of the T cell receptor does result in desensitization of SPR. We are currently investigating whether some heterologous desensitization mechanism is taking place and whether a particular kinase is involved.

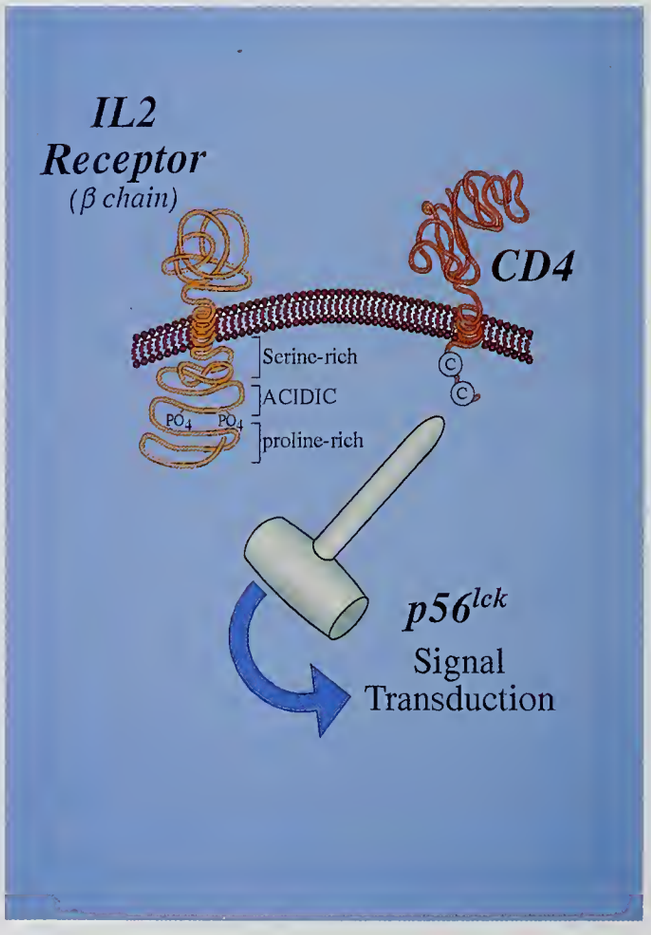
In collaboration with Phyllis Gardner at Stanford University, we have also been able to demonstrate, using patch-clamp techniques, that activation of the SPR in these transfected lymphocytes results in the opening of a chloride channel. In addition, these changes can be nullified by injecting the cells with a calcium/calmodulin kinase inhibitor peptide.

Our early observations that stimulation of these cell lines by substance P resulted in both IP_3 and cyclic AMP activation have now been extended. Looking at the cells' respective nuclear regulatory elements, we have demonstrated that substance P stimulation results in increased expression of the AP-1– and the cAMP-responsive elements in these cells. Furthermore, using Western and Northern blotting techniques, we have also been able to demonstrate that the proto-oncogenes *fos* and *jun* are also up-regulated following substance P stimulation. Consequently, signal transduction analysis suggests that a dual activation signal pathway may be activated when the SPR is expressed in lymphoid cells.

We have studied the functional consequences of stimulating these cells with the peptide substance P by examining the expression of a number of cell surface antigens and their modulation on Jurkat SPRs. We find that in Jurkat SPR-positive cells, substance P stimulation results in a down-regulation of the LFA-1 (integrin) surface antigen and an up-regulation of the CD2 surface antigen. In addition, when combined with the mitogen phytohemagglutinin, substance P stimulation results in an increased expression of the interleukin-2 receptor. We continue to examine these results in order to delineate further how tachykinin peptides may modulate immune responses.

Joseph Fisher and Sandra Biroc have now begun an extensive *in situ* hybridization study of the expression of agrin and alternate agrin transcripts in the developing rat. Preliminary results suggest that agrin is extensively expressed in unique sites within the developing brain. Moreover, Dr. Fisher has now expressed full-length recombinant agrin in a number of cell types and is examining the molecule's biochemical properties. In particular, the amino terminus of the molecule

has some unique sequences characterized by protease inhibitor domains. He has begun a study of the molecule's protease inhibitor activity to see if it plays a role during the remodeling process in neuronal development. We expect the studies regarding receptor signal transduction mechanisms and early developmental remodeling processes in the nervous system to help us further understand how the immune and nervous systems develop and interact.



Cell surface receptors on lymphocytes sometimes share intracellular signaling mechanisms. In this case, the receptor for a T cell growth factor (interleukin-2 or IL-2) and the protein molecule CD4, which assists these cells in the recognition of foreign antigens, both interact with a lymphocyte-specific protein-tyrosine kinase, $p56^{lck}$—a potent signaling molecule that induces proliferation. A highly acidic region of the IL-2 receptor's β-chain mediates this interaction, while a very different domain in the CD4 protein, containing two critical cysteine residues (c), performs a similar function for this receptor.

Diagram from Roger Perlmutter, based on work in his laboratory.

Molecular Basis of Lymphocyte Signaling

Roger M. Perlmutter, M.D., Ph.D.—Investigator

Dr. Perlmutter is also Professor of Immunology, Medicine, and Biochemistry at the University of Washington School of Medicine. He received his B.A. degree from Reed College and his M.D. and Ph.D. degrees from Washington University, St. Louis, where he studied with Joseph Davie. After clinical training in internal medicine at Massachusetts General Hospital, Boston, and the University of California, San Francisco, he became Senior Research Fellow and later Instructor in Biology at the California Institute of Technology, where he worked with Lee Hood.

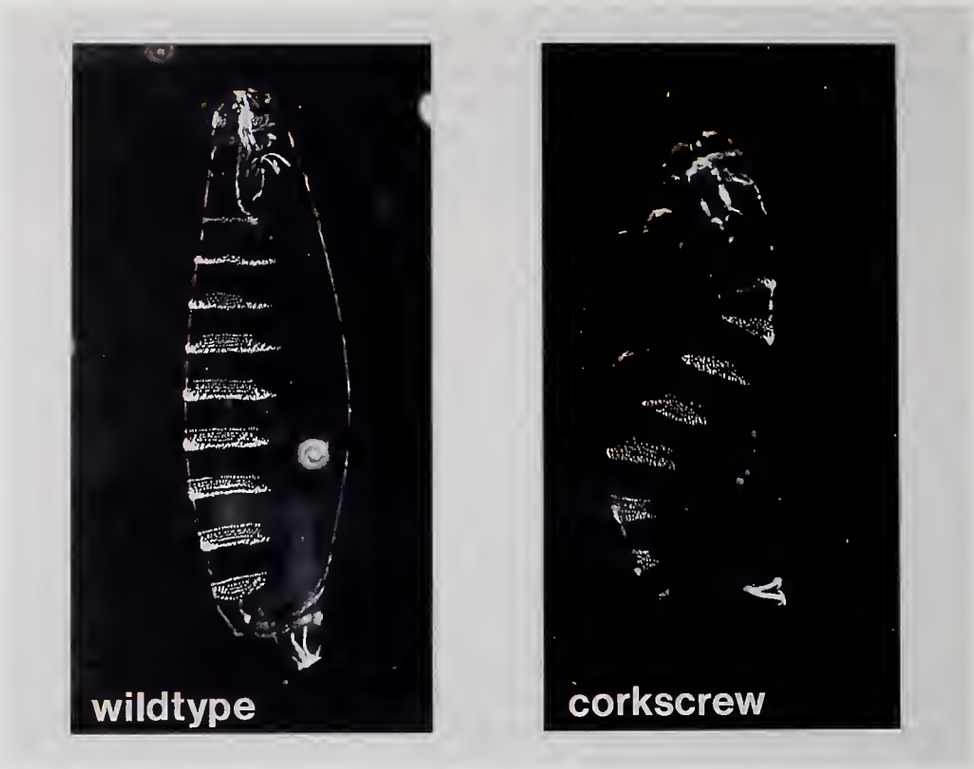
IMMUNE recognition of potentially injurious foreign macromolecules requires the elaboration of an enormous repertoire of clonally restricted receptors (antigen receptors) on the surfaces of lymphoid cells. These receptors are sufficiently heterogeneous to permit recognition of virtually the entire universe of infectious organisms. Interaction of these receptors with cognate antigen provokes a stereotyped response leading to cell proliferation and the production of soluble mediators of inflammation. During the past several years, analysis of the mechanisms responsible for antigen receptor diversification has stimulated interest in a related question: How is the signal from a lymphocyte antigen receptor transmitted to the cell interior? Our laboratory has adopted a molecular genetic approach to the dissection of signaling pathways in immune cells.

Initially we identified a lymphocyte-specific enzyme, similar in structure to proteins known to transmit growth-promoting signals in nonlymphoid cells, that modifies the behavior of target proteins by catalyzing the addition of phosphate groups onto certain tyrosine amino acids in substrate proteins. The gene that encodes this lymphocyte-specific kinase was identified by virtue of its overexpression in a murine lymphoid malignancy. Moreover, we were able to demonstrate that a single point mutation in this *lck* gene enabled it to confer malignant properties on cells maintained in culture. Thus the *lck* gene encodes a protein that is capable of altering the growth properties of at least some cell types. Since *lck* is normally expressed only in lymphocytes, there is reason to believe that its product assists in regulating lymphocyte proliferation.

Biochemical studies support this view. In particular, we and others have recently demonstrated that the *lck*-encoded kinase is physically associated with proteins that form part of the antigen receptor on T lymphocytes. Our studies also enabled us to identify two additional protein-tyrosine kinases that are specifically expressed in immune cells. In each case there is reason to believe that the kinase is physically coupled to a cell surface receptor involved in immune recognition.

To investigate the functional importance of these protein-tyrosine kinase signaling elements, we have developed methods for manipulating the expression of each gene in its appropriate cellular context. Using this approach, we have begun to dissect the hierarchy of signal transduction events precipitated by normal immune recognition. For example, we found that overexpression of an activated *lck* gene leads to extraordinarily rapid development of thymic tumors in mice. Hence altered expression of the *lck* gene can directly affect lymphocyte proliferation. In a related series of experiments we learned that inhibition of the function of the *lck* gene completely disrupts normal mechanisms that permit development of T lymphocytes. In fact the level of *lck*-encoded protein must be maintained within a very narrow range. Even very modest (twofold) changes in the abundance of this kinase are incompatible with normal T lymphocyte development. These observations probably reflect the fact that the *lck*-encoded kinase participates in a large number of receptor signaling pathways, including some pathways that are activated by lymphocyte-specific growth factors such as interleukin-2.

Disturbances in lymphocyte signaling almost certainly contribute to the pathogenesis of lymphoproliferative and immunodeficiency diseases in humans. A detailed understanding of immune cell signaling mechanisms should permit the design of more-effective therapeutic strategies for the treatment of immune system dysfunction.



Darkfield preparation of a wild-type and a corkscrew *embryo. The twist in the mutant embryo is due to a perturbed establishment of cell fates at the termini of the blastoderm.*

From Perkins, L.A., Larsen, I., and Perrimon, N. 1992. Cell *70:225–236. Copyright © 1992 by Cell Press.*

Genetic Dissection of a Signal Transduction Pathway in Drosophila melanogaster

Norbert Perrimon, Ph.D.—Assistant Investigator

Dr. Perrimon is also Assistant Professor of Genetics at Harvard Medical School. Of French nationality, he was educated at the University of Paris VI, where he majored in biochemistry. His thesis, with Madeleine Gans as advisor, was on Drosophila *genetics. He moved to Case Western Reserve University as a postdoctoral research fellow with Anthony Mahowald. He became a Lucille P. Markey Scholar in Biomedical Sciences while in Cleveland. He then assumed his present position at Harvard Medical School.*

INTERCELLULAR communication is a major player in the establishment of developmental patterns. For example, in the early *Drosophila* embryo, determination of cell fates at the termini requires the normal activities of genes from two different cell types, the maternal follicle cells and the oocyte. The current model is that the transmembrane tyrosine kinase receptor encoded by the gene *torso* is locally activated at the egg termini by cues emanating from the follicle cells. This localized activation of *torso* is believed to trigger a phosphorylation cascade in the egg which ultimately controls the expression of the transcription factors *tailless* and *huckebein*. Knocking out this signaling pathway has detrimental effects on embryonic development. Since cell fates at both termini are perturbed, the resulting embryos lack most head and all tail structures.

Our laboratory has focused on identifying the genetic components involved in transduction of the signal from torso, the membrane-bound tyrosine kinase, to the nucleus. Thus far we have characterized two genes, *l(1)pole hole* and *l(1)corkscrew*, that are involved in this process. Genetic epistasis experiments have demonstrated that both these genes act downstream of the *torso* protein activity. Furthermore, *l(1)corkscrew* acts by up-regulating the activity of *l(1)pole hole*.

We previously showed that the *l(1)pole hole* gene product is the homologue of the mammalian *Raf-1* proto-oncogene and encodes a serine/threonine kinase. Recently we discovered that *l(1)corkscrew* encodes a protein-tyrosine phosphatase similar to the mammalian PTP1C protein. Identification of the *l(1)pole hole* and *l(1)corkscrew* gene products has strengthened the current model that torso signaling involves a phosphorylation cascade, since both genes encode proteins that have the ability to affect the level of phosphorylation of intracellular components.

To identify additional molecules involved in torso signaling, we have taken a genetic approach. Screens for second-site suppressors and enhancers of *l(1)pole hole* and *l(1)corkscrew* mutations have successfully identified a number of loci involved in the torso signaling pathway. Future work will involve a detailed characterization of these suppressors.

In addition to the genetic approach described above, we are utilizing a biochemical screen to identify and characterize other components of this signal transduction pathway. Involved is the cloning of genes encoding proteins that bind directly to activated receptor tyrosine kinase. In this way we are isolating proteins that respond directly to the *torso*-encoded receptor.

Establishment of cell fate at the termini of the embryo provides a unique genetic system to dissect the cascade by which activation of a receptor tyrosine kinase controls the expression of transcription factors. A combination of classical genetics and molecular and biochemical techniques will allow characterization of the components involved in the various steps of receptor tyrosine kinase signaling.

In addition, the homology between *torso* and the mammalian gene for platelet-derived growth factor (PDGF), between *l(1)pole hole* and the mammalian *Raf-1* proto-oncogene, and between *l(1)corkscrew* and the human PTP1C enzyme suggests that biochemically this signal transduction pathway may have been conserved in evolution between organisms as diverse as *Drosophila* and humans. Thus characterization of this pathway in *Drosophila* may help elucidate the functions of the homologous mammalian proteins.

Gene Regulation and Immunodeficiency

B. Matija Peterlin, M.D.—Associate Investigator

Dr. Peterlin is also Associate Professor of Medicine and of Microbiology and Immunology at the University of California, San Francisco. He obtained his undergraduate degree in chemistry and physics at Duke University and his M.D. degree from Harvard Medical School. His postdoctoral work was performed with Jacob Maizel and Philip Leder at NIH and with Hugh McDevitt at Stanford University. As a rheumatology fellow at Stanford, he chanced upon a family with the bare lymphocyte syndrome, which stimulated his interest in this genetic disorder. He is a member of the American Society for Clinical Investigation.

SOME years ago we described a variant of the genetic disorder called the bare lymphocyte syndrome (BLS), in which the patient's lymphocytes fail to express either class I, class II, or both major histocompatibility determinants on their cell surfaces. These transplantation antigens are essential for the development of the immune system, for tumor surveillance, for eradication of viral infections, and for normal immune responses. Thus it is not surprising that BLS patients are severely immunocompromised, fail to make antibodies, or have autoimmune diseases. In addition, this autosomal recessive syndrome is one of only two known inherited deficiencies of a regulatory gene in humans.

By fusing defective cells from different patients and those obtained by mutagenesis in tissue culture, four genetic complementation groups of BLS were found. The isolation of their defective genes should make possible prenatal diagnoses through use of specific genetic probes and possibly lead to the cure of BLS by the targeting of normal genes into the bone marrow of affected patients.

To study the defective gene in BLS, we first examined regions that regulate B cell–specific and interferon-γ (IFN-γ)-inducible expression of class II genes. Next, we looked at proteins that bind to these DNA sequences and compared class II–specific factors in various cell types. Distinct patterns of DNA-binding proteins were found in B cells, IFN-γ–inducible cells, and T cells.

We cloned several cDNAs that code for proteins that bind to B cell–specific and IFN-γ–inducible sequences in class II promoters. One cDNA codes for Jun, which forms active Jun/Fos heterodimers in cells that do not express class II determinants. Of the remaining two cDNAs, one codes for a B cell–specific helix-loop-helix protein and the other for an ETS-like protein. We are currently studying their genetic organization, expression, structure, and function. By expressing one of these full-length cDNAs in human cells, we hope to rescue class II gene expression in one type of BLS. In parallel with direct biochemical studies, we are also using genetic approaches to rescue regulatory defects in BLS.

In setting up these genetic approaches, we first tested a well-known viral trans-regulatory system —namely trans-activation of the human immunodeficiency virus (HIV) by the virally encoded Tat protein. The precise mechanism of Tat action had not been defined. We discovered that Tat acts slightly downstream from the promoter to modify HIV transcription so that efficient copying of the viral genome ensues. Factors assembled near the site of initiation of HIV transcription tether the transcription complex to the promoter. The addition of Tat, which binds to an RNA stem-loop in the process of nascent transcription, releases this transcription complex. Efficient elongation of transcription and clearance of the promoter follow. New transcription complexes can then assemble, interact with Tat, and move quickly through the viral genome. Interactions between Tat, the RNA stem-loop, and cellular proteins have been defined. For example, using a heterologous RNA-tethering mechanism (that of the coat protein of bacteriophage R17 that binds to its operator), we mapped activation and RNA-binding domains of Tat. By studying rodent cells and somatic cell hybrids between rodent and human cells, we defined a cellular RNA-binding complex that facilitates interactions between Tat and TAR and is encoded on human chromosome 12. We hope that interfering with trans-activation by Tat will lead to new therapies for AIDS (acquired immune deficiency syndrome) and AIDS-related disorders.

Since upstream promoter sequences are also essential for HIV replication, we clarified interactions between host cell factors and viral sequences (long terminal repeat, LTR). Increased rates of initiation of HIV transcription were observed in activated T cells and macrophages. These result from actions of nuclear proteins that are also required for T cell and macrophage effector functions and for T cell proliferation.

Differences between LTRs of HIV types 1 and 2 were observed that might explain the longer latency and attenuated clinical course of HIV-2 infection. Furthermore, effects of trans-activators encoded by several DNA viruses on HIV transcrip-

tion were examined. Since effects of these agonists and Tat were synergistic, infection by DNA viruses might be an important cofactor in progression from latent disease to clinical AIDS.

Thus the quest to rescue the expression of class II major histocompatibility genes in a rare human congenital disease has led to genetic approaches to the study of transcription by RNA polymerase II and to the elucidation of a potent new transcriptional regulatory mechanism.

Mechanism of Action of Polypeptide Growth Factors

Linda J. Pike, Ph.D.—Associate Investigator

Dr. Pike is also Associate Professor in the Department of Biochemistry and Molecular Biophysics at Washington University School of Medicine, St. Louis. She received her B.S. degree in chemistry from the University of Delaware and her Ph.D. degree in biochemistry from Duke University, where she studied with Robert Lefkowitz. Her postdoctoral training was done in the laboratory of Edwin Krebs at the University of Washington, Seattle.

A number of low-molecular-weight polypeptides have been shown to regulate cell growth. These growth factors bind to specific receptors on the surface of cells. Through a complex series of reactions, the binding of the growth factor to its receptor stimulates the cell to grow and divide. Until recently, little was known of the mechanism by which growth factors induce cell division. It is now recognized, however, that the receptors not only bind the appropriate growth factor but also have an enzymatic activity.

The receptors catalyze the transfer of a phosphate group from adenosine triphosphate to tyrosine residues in selected protein substrates. The enzyme possessing this activity is called a tyrosine protein kinase. Typically the phosphorylation of a protein by a kinase leads to changes in the activity of the protein. Although much is known about the growth factor receptor kinases, the substrates for these enzymes have not been identified.

My laboratory is involved in studies of the mechanism by which the binding of epidermal growth factor (EGF) to the outside of the cell elicits a biological response inside. This is referred to as signal transduction. Because they are extremely responsive to EGF, the A431 line of human epidermal carcinoma cells is used as the model system in most of our studies.

Phosphatidylinositol Metabolism

One of the earliest responses of A431 cells to EGF is an increase in the metabolism of a particular phospholipid, phosphatidylinositol. This important lipid serves as a precursor for the generation of two intracellular compounds that activate various enzymes and thereby mediate the effects of EGF within the cell. One of the enzymes involved in phosphatidylinositol metabolism is phosphatidylinositol kinase. Since the activity of this enzyme is stimulated by EGF, it represents a potential substrate for phosphorylation by the EGF receptor tyrosine protein kinase.

We have purified the phosphatidylinositol kinase from both A431 cells and human placenta. This 55-kDa enzyme is active as a monomer—that is, as a single polypeptide chain. It phosphorylates phosphatidylinositol on the 4 position of the inositol ring and hence is distinct from another phosphatidylinositol kinase that phosphorylates the ring on the 3 position. Information regarding the sequence of the amino acids that make up the phosphatidylinositol 4-kinase (PI 4-kinase) was obtained. Molecular biology techniques were used to isolate and characterize the cDNA encoding the PI 4-kinase. From this cDNA the amino acid sequence of the entire PI 4-kinase was deduced and was found to be unique. The sequence showed limited similarity to other proteins that bind inositol phosphate or carry out the phosphorylation of other sugars, suggesting that the PI 4-kinase may be derived from enzymes that metabolize sugars. Analysis of the types of mRNAs present in a variety of tissues that encode the PI 4-kinase indicated that at least two types of mRNA are present and that the amount of message is regulated by physiological stimuli. Efforts are now directed toward producing a cell line in which this PI 4-kinase cDNA is overexpressed, leading to elevated levels of PI 4-kinase activity within the cell. The effects of this overexpression on the growth properties of the cells will be investigated. This work is supported by a grant from the National Institutes of Health.

Another enzyme involved in phosphatidylinositol metabolism is a phosphatidylinositol monophosphate phosphatase. This enzyme catalyzes the reverse of the reaction catalyzed by the PI 4-kinase—that is, it removes the phosphate from the 4 position of the inositol ring of phosphatidylinositol monophosphate. Although this enzyme has not been studied previously, its position in the metabolic pathway of phosphatidylinositol suggests that it may be important in the overall regulation of the pathway. We have characterized it with respect to its kinetic properties, substrate specificity, and response to various inhibitors. We have purified the enzyme to a high degree and have shown that it is a 95-kDa glycoprotein. Unlike other enzymes in this pathway, it appears to have an extracellular domain. The possibility that its activity is regulated through the binding of an extracellular mediator is being investigated.

It has recently been postulated that the inositol phospholipids may function in the regulation of cell shape via interactions with proteins that control the polymerization of actin. Using EGF and A431 cells, we have shown that there is no correlation between EGF-induced changes in cell shape and changes in the levels of the classical phosphoinositides, phosphatidylinositol 4-phosphate and phosphatidylinositol 4,5-bisphosphate. Further experiments have also failed to document the involvement of novel inositol phospholipids, phosphorylated on the 3 position of the inositol ring, in the regulation of cell shape by EGF. These data suggest that EGF may control the actin cytoskeleton simply by a mechanism that involves phosphorylation of cytoskeletal proteins rather than indirectly by causing alterations in inositol phospholipid levels.

Desensitization of the EGF Receptor

When A431 cells are treated with large doses of EGF, washed, and subsequently rechallenged with EGF, they fail to respond to the growth factor. This phenomenon is known as desensitization. Our studies have shown that when the EGF receptor becomes desensitized, it is no longer internalized into the cells, and EGF no longer stimulates phosphatidylinositol metabolism. This EGF-induced desensitization is specific for the EGF receptor, because the responsiveness of other receptors is not decreased after EGF treatment.

The EGF receptor itself is a monomeric protein, a single chain. Upon binding of EGF to its receptor, two of the receptor monomers come together to form an EGF receptor dimer. This dimer is the form that is active in signal transduction. We have shown that desensitized EGF receptors do not transduce a signal because they cannot undergo this EGF-induced dimerization.

Our data suggest that the desensitization of the EGF receptor results from its phosphorylation by a protein kinase. We have identified a protein kinase in A431 cell cytosol that is activated by EGF and appears to be involved in receptor desensitization. The kinase catalyzes the phosphorylation of the EGF receptor *in vitro*. Consistent with what has been observed in whole cells, phosphorylation of EGF receptor monomers by this kinase leads *in vitro* to an inhibition of the ability of the phosphorylated monomers to dimerize.

The kinase phosphorylates the EGF receptor on a serine residue in the second half of the receptor molecule. Using techniques of molecular biology, we have altered this site and are in the process of characterizing the properties of cells expressing this mutated form of the EGF receptor.

Protein Structures, Molecular Recognitions, and Functions

Florante A. Quiocho, Ph.D.—Investigator

Dr. Quiocho is also Professor of Biochemistry and Structural Biology and of Molecular Physiology and Biophysics at Baylor College of Medicine. He obtained his Ph.D. degree in biochemistry at Yale University and then did postdoctoral research in chemistry at Harvard University. He was a member of the Rice University faculty before joining the Baylor faculty. Dr. Quiocho has been a visiting research scientist at Oxford University, a research fellow of the European Molecular Biology Organization (EMBO), and a Guggenheim fellow.

LIGAND specificity and the activity of proteins are derived from their precise three-dimensional structures. Using mainly x-ray crystallographic techniques, our laboratory is engaged in atomic-level elucidation of the structures and functions of several proteins (including enzymes) involved in biologically important processes. To complement our work, we also employ biochemical and recombinant DNA techniques.

Adenosine Deaminase

Adenosine deaminase (ADA) is one of the major enzymes in purine metabolism, catalyzing the irreversible hydrolysis of adenosine or deoxyadenosine to the respective inosine product and ammonia. The enzyme is found in nearly all mammalian cells and plays a central role in maintaining competency of the immune system, among several other functions. Lack or deficiency of ADA is associated with severe combined immunodeficiency disease (SCID), a genetically inherited disorder usually fatal within two years of birth if left untreated.

Last year we determined the crystal structure of ADA with bound 6-*R*-hydroxyl-1,6-dihydropurine ribonucleoside (HDPR), a nearly ideal transition-state analogue inhibitor. We have since elucidated the structures of complexes of the enzyme with the following ligands: 1-deazaadenosine, a substrate analogue; 2′-*R*-deoxycoformycin, a potent transition-state analogue and a chemotherapeutic agent for the treatment of hairy cell leukemia; and inosine, the product of the deamination of adenosine. All these structures have provided us with a molecular anatomy of the various steps associated with ADA's catalytic activity.

All of the four crystal structures of the deaminase indicated above have been determined and refined at pH 4.2, where the enzyme is only 20 percent active. We have also carried out the refinement of the structure of ADA complexed with HDPR at pH 6, where the enzyme is fully active. The structures at either pH are essentially the same.

Antibody-Antigen Interactions

Because monoclonal antibodies against extracellular polysaccharide antigens exhibit very stringent specificity, they have been used in blood-grouping and in differentiating bacterial serogroups and serotypes. We previously obtained crystals of the Fab fragment of the antibody raised against the surface polysaccharide O-antigen of *Shigella flexneri.* In the past year we determined and refined the structures of the Fab and its complexes with a trisaccharide, α-Rha(1-3)α-Rha(1-3)β-GlcNAc, and a pentasaccharide, α-Rha(1-2)α-Rha(1-3)α-Rha(1-3)β-GlcNAc-(1-2)α-Rha. Both oligosaccharides contain determinants of the O-antigen serotype of the bacteria.

This structural work is in line with our interest in protein-carbohydrate interactions. Moreover it is relevant to clinical problems, as oligosaccharide epitopes of bacterial and tumor cell surfaces are considered to be disease markers and targets for therapeutic antibodies.

Aldose Reductase

Aldose reductase catalyzes the NADPH-dependent reduction of a wide variety of carbonyl-containing compounds to their corresponding alcohols, with a broad range of catalytic efficiencies. Steroids are the best substrates and sugars the least favorable. Although the enzyme is found in a variety of cells, its physiological function has not been firmly established. A role in reducing the hyperglycemia of diabetes mellitus has been linked to diabetic complications affecting the lens, retina, peripheral nerves, and kidney. Drugs designed to control these complications have not been clinically successful to date because of lack of specificity or inefficacy.

In collaboration with Kurt Bohren and Kenneth Gabbay of the Baylor College of Medicine, we have obtained excellent diffracting crystals of recombinant aldose reductase (from human placenta) with bound NADPH. We have determined and refined the enzyme's three-dimensional, 1.65-Å resolution structure. The enzyme has a parallel β/α-barrel motif, with eight central β-strands connected by eight peripheral α-helices.

This establishes a new motif for NADP/NAD-binding oxidoreductases and the first structure of the superfamily of aldo-keto reductases.

The substrate-binding site is located in an extremely hydrophobic elliptical pocket at the carboxyl-terminal end of the β-barrel. The nicotinamide group of the NADPH is at the bottom of the deep pocket. Although the hydrophobic nature of the active site greatly favors aromatic (e.g., steroids) and apolar substrates, it is not well suited for binding of highly polar monosaccharides, which are believed to figure in the pathogenesis of diabetic complications. The determination of the structure of aldose reductase paves the way for rational design of specific inhibitors that might provide molecular understanding of the catalytic mechanism, as well as possible therapeutic agents for the prevention of diabetic complications.

Periplasmic Receptors for Active Transport and Chemotaxis

The family of binding proteins that serve as initial periplasmic receptors for bacterial active transport and chemotaxis continues to be a gold mine for detailed study of protein structure and molecular recognition of a variety of ligands. We continue to push the structure refinements of the seven different periplasmic receptors to much higher resolutions—the sulfate-binding protein (at 1.7-Å resolution), phosphate-binding protein (1.17 Å), L-arabinose–binding protein (1.7 Å), D-galactose/D-glucose–binding protein (1.5 Å), maltodextrin-binding protein (1.7 Å), leucine/isoleucine/valine-binding protein (1.7 Å), and leucine-specific–binding protein (1.7 Å).

Electrostatic Interactions in Molecular Recognition of Ligands

Electrostatic interactions are among the key factors determining the structure and function of proteins. The refined structure of the liganded form of sulfate-binding protein shows that the bound sulfate dianion is completely buried and bound by hydrogen bonds and van der Waals contacts. The bound sulfate is adjacent to the amino termini of three helices and is coupled via a peptide unit to a positively charged His residue. Nevertheless, using site-directed mutagenesis and theoretical analysis, we have shown that helix macrodipoles and the His residue contribute almost nothing to ligand fixation and charge compensation. It is the collection of local dipoles immediately surrounding the sulfate that is responsible for charge compensation.

Molecular Approaches to Olfaction

Randall R. Reed, Ph.D.—Associate Investigator

Dr. Reed is also Associate Professor in the Departments of Molecular Biology and Genetics and of Neuroscience at the Johns Hopkins University School of Medicine. He received his bachelor's degree in biophysics from Johns Hopkins and his Ph.D. degree from Yale University. His postdoctoral research was done with Philip Leder at Harvard Medical School.

OLFACTION is among the oldest of the sensory systems. All multicellular and many unicellular organisms have evolved sensitive chemosensory systems able to detect and identify natural chemical substances. The olfactory system of vertebrates and analogous systems for the other senses—vision, hearing, taste, and touch —allow the conversion of external stimuli into nerve impulses. In mammals, the olfactory system is exquisitely sensitive, capable of detecting some odorants present at a concentration of only a few parts per trillion.

The ability of the olfactory system to discriminate thousands of different odorants suggests a complex coding mechanism. However, the biological basis for this coding remains a mystery. Unlike the visual and auditory systems, which need only encode information on frequency (or wavelength) and intensity, the olfactory sense requires multidimensional information. These considerations suggest a complex signal transductory process.

The signal transductory pathway for olfaction can be divided temporally and spatially into several distinct steps. The first consists of the solubilization and concentration of airborne odorants. Considerable experimental data suggest that components of the mucus are able to concentrate odorants several thousandfold. We previously identified cDNA clones encoding proteins that are present at high concentration in the mucus and appear to bind odorants. These proteins, from rat and frog, are members of the retinol-binding protein family, many of which have been shown to solubilize hydrophobic ligands in serum. They are likely to play a similar role in the olfactory system.

The recognition of the chemical structure of an odorant and transduction of that information across the plasma membrane is a poorly understood process. Some investigators hypothesize that odorants interact directly with the lipid membrane, but it is difficult to see how the ability to discriminate stereoisomeric compounds could be accommodated by such a system. Several years ago we began to test an alternative hypothesis: that binding of odorants to specific membrane-associated proteins leads to intracellular changes in the primary sensory neuron.

The detection of odorant-stimulated activation of second messengers in olfactory neurons suggested an analogy to sensory systems. In visual transduction, which is the best characterized of these systems, sensitivity is achieved through a second messenger cascade consisting of the membrane-bound receptor rhodopsin, a rod photoreceptor-specific GTP-binding protein, transducin, and a cyclic GMP phosphodiesterase. Light-stimulated decrease in the concentration of intracellular second messenger leads to modulation of the cellular membrane. Activation of adenylyl cyclase in preparations of rat olfactory cilia depends on the presence of guanine nucleotides. The observation of odorant-stimulated GTP-dependent adenylyl cyclase activity argues strongly for a similar GTP-binding protein–coupled signal transductory pathway in olfaction.

We have identified a GTP-binding protein as well as an adenylyl cyclase expressed exclusively in olfactory sensory neurons. Moreover, these components are localized to the olfactory cilia, where the initial events in olfactory signal transduction are thought to occur. The G_{olf} protein, which is highly homologous to a GTP-binding protein that stimulates adenylyl cyclase in other systems, is expected to interact directly with the olfactory receptors. Additionally, we have demonstrated that G_{olf} can couple receptor activation to increases in intracellular cAMP in cell lines deficient for the stimulatory G protein, G_s.

Identification of genes encoding olfactory receptors may reveal how these structures are able to detect thousands of different odors. Recently, Linda Buck and Richard Axel (HHMI, Columbia University College of Physicians and Surgeons) have described a gene family that encodes proteins expressed in olfactory tissue. We have identified additional members of this large family, utilizing the polymerase chain reaction technique. These experiments reveal that expression of the mRNA that encodes these putative olfactory receptors is confined to the sensory neurons of the olfactory epithelium.

We are now engaged in characterizing these receptors biochemically and in screening for the specific ligands that activate them. The subcellular localization of the receptor proteins is being examined by means of antibodies directed against conserved regions. These molecular tools should allow us to address some important questions. Does each of the several million olfactory sensory cells express a single receptor protein species? How are the genes that encode these receptors organized, and how is their expression regulated?

The final step in the transduction of odorant stimuli is the generation of the intracellular signal and the firing of an action potential. Specialized forms of second messenger–generating enzymes and novel ion channels play important roles in this process. We have identified cDNA clones encoding three distinct forms of adenylyl cyclase and are investigating the regulation of this important enzyme in olfactory tissue. Olfactory neuronal adenylyl cyclase type III has biochemical properties that would be advantageous for an enzyme involved in sensory transduction. Electrophysiologic experiments have identified cyclic nucleotide–gated ion channels in olfactory neurons, and we have recently isolated and characterized cDNA clones from olfactory tissue that encode ion channels with properties similar to those found in the visual system.

The olfactory system is also interesting as a model for neuron differentiation and development. The olfactory neuroepithelium is the only neuronal tissue in adult mammals that undergoes continual regeneration. The lifetime of sensory neurons is approximately 40 days, after which they are shed from the epithelium and replaced from a population of neuroblast-like precursor cells. Moreover, if the nerve leading from the sensory neurons to the olfactory bulb is severed, all 10 million receptor cells are rapidly lost and subsequently replaced in a relatively synchronous fashion. My laboratory is beginning to address the mechanism of regulation of olfactory neuron–specific genes.

We have identified a specific DNA sequence upstream of all the genes known to be expressed specifically by the olfactory neurons and have observed in olfactory homogenates a binding activity specific for that sequence. This putative transcriptional regulator might direct the expression of the entire repertoire of genes involved in olfactory signal transduction and neuronal maturation.

We have recently identified a number of cDNA clones that encode proteins expressed only in the mature sensory neurons and are attempting to elucidate their role in the cell. A novel group of proteins, those expressed transiently during neuron maturation, may include receptors for neurogenic as well as neurotrophic factors. Several of the genes we have identified appear to encode membrane-bound or cell surface proteins.

We are continuing to use several techniques to elucidate the mechanism of signal transduction. Likewise, the identification of proteins associated with the replacement of olfactory neurons provides the tools to study neural development, not just in the olfactory system but also in other areas of the brain. In the future we will focus on the molecular components that underlie the complex mechanisms of signal transduction, signal processing, and the formation of neural connections.

The Molecular Basis of Hereditary Diseases of the Kidney

Stephen T. Reeders, M.D.—Associate Investigator
Dr. Reeders is also Associate Professor of Internal Medicine and Genetics at Yale University School of Medicine. He attended Cambridge University with the intention of majoring in physics, but, realizing that developments in molecular biology were providing the basis for new approaches to the study of human disease, he switched to medicine and continued to study at Oxford University. After qualifying in medicine, he sought clinical training in intensive care, cardiology, nephrology, and neurology. Then, with Sir David Weatherall at Oxford, he began to use molecular genetic techniques to study human disease, with emphasis on hereditary diseases of the kidney, diseases that heretofore had received little attention from geneticists.

CRITICAL for normal functioning of the kidney is the integrity of the glomerular basement membrane (GBM), a complex extracellular structure that forms one of the main barriers between the blood and the urine. The GBM is composed of several proteins, including five related but subtly different collagens that interact to form a chicken-wire mesh holding the membrane together. One of the interests of our laboratory is Goodpasture syndrome, an autoimmune disorder in which, for unknown reasons, autoantibodies are suddenly targeted at the collagen components of basement membrane in the lungs and kidneys. In the kidney, these autoantibodies produce a sudden and devastating severe inflammation, which frequently leads to acute renal failure, irreversible unless treated. The nephritis is often accompanied by autoimmune lung damage, manifested by bleeding into the alveoli.

Previous studies have shown that the probable target of Goodpasture autoantibodies is the $\alpha 3$ chain of basement membrane collagen. To understand the pathogenesis more clearly, we undertook to isolate and purify the collagen chain to study its structure. Because the protein is present in very small amounts and is accompanied by four similar proteins, purifying it has proved difficult. We have therefore isolated, cloned, and sequenced the gene for the $\alpha 3$ chain of basement membrane collagen and have used the sequence information to predict the primary structure and compare this protein with other basement membrane collagens.

In collaboration with Billy Hudson (Kansas City), we used knowledge of the primary structure to identify several potential antibody-binding sites (epitopes) in the $\alpha 3$ molecule. We synthesized short peptides and used them to test the binding of some of these sites, which we localized to within a small region at the carboxyl terminus. At least one of the peptides has very high affinity for Goodpasture antibodies and adsorbs them from patients' serum. Knowledge of the epitope structure should enable us to develop a means of selectively adsorbing Goodpasture antibodies, opening possibilities for a new treatment modality. In addition, this information may provide clues to the development of autoimmunity in this disorder.

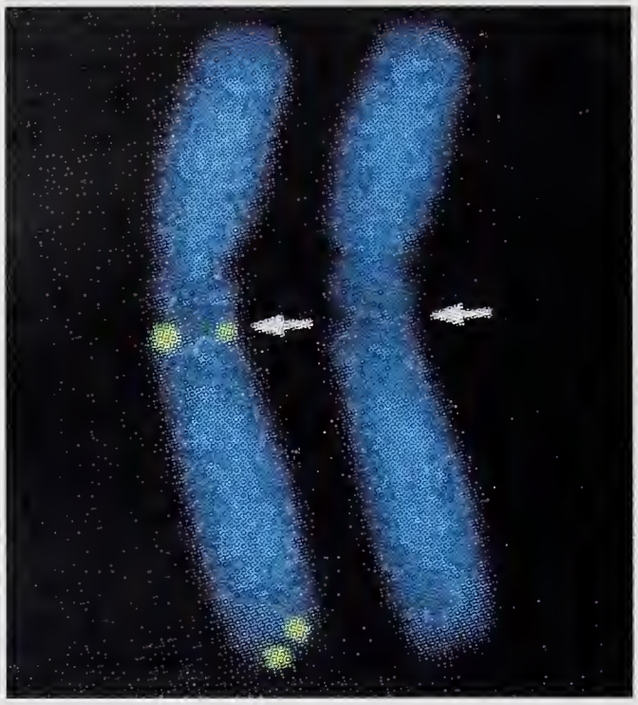
Mariko Mariyama, a Howard Hughes associate in our laboratory, has recently isolated a novel basement membrane collagen, $\alpha 4$. The comparison of the structure of this molecule with the known basement membrane collagen allows the evolution of these molecules to be inferred. Two classes of the molecule exist in all species from roundworm to humans, suggesting a functional divergence. The expression of $\alpha 3$ and $\alpha 4$ occurs in a limited number of tissues, and they are regionally localized within these tissues. In kidney, for example, they are found in the glomerulus but not in the tubular or vascular endothelial basement membranes. Our gene mapping data suggest that the expression of $\alpha 3$ and $\alpha 4$ is coordinated by transcription of the genes from opposite DNA strands. A similar arrangement has been found for the $\alpha 1$ and $\alpha 2$ genes, homologues of $\alpha 1$ and $\alpha 2$.

One of the major projects in our laboratory is a study of the molecular and cellular pathology of autosomal dominant polycystic kidney disease (ADPKD), one of the commonest causes of kidney failure in humans, affecting at least 1 in 1,000 of the population. The disease is an enormous burden to families and the community, since the majority of patients develop irreversible kidney failure in middle life and require dialysis or transplantation for survival.

Having previously ascertained that the majority of the inherited mutations in ADPKD lie close to the tip of the short arm of chromosome 16, we have isolated and cloned a small segment of DNA (550,000 base pairs) that includes the mutated gene. This region turns out to be extremely gene rich, and we have already isolated 22 genes from within it. These include novel cyclin A–like and *ras*-like genes, a gene encoding a zinc finger protein, and a gene having homology to the β-

subunit of the G protein family. Since we have not been able to detect any large-scale deletions or rearrangements affecting any of the 22, we have begun to examine the sequence of these genes in detail and to look for mutations that may affect only one or two nucleotides. This work is supported by a grant from the National Institutes of Health.

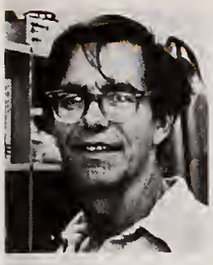
Our laboratory has been interested in the structure of telomeres, the ends of chromosomes. We have shown that banks of repetitive sequence reminiscent of the sequence of human telomeres (TTAGGG) are also present at other sites within the human genome. One of the most interesting arrays of telomere-like sequence is embedded in the middle of the long arm of chromosome 2. Comparison of the chromosome banding pattern of humans with that of several closely related apes suggests that this region of the chromosome contains a point at which two ancestral ape chromosomes fused. Jaap IJdo, a Howard Hughes associate, has cloned this fusion point and shown that it consists of a head-to-head telomere-telomere fusion.



In situ *hybridization of a cloned DNA segment containing subtelomeric sequences to a human chromosome 2 (left). The fluorescent probe recognizes the sequences at the tips of several chromosomes, including those of the long arm of chromosome 2 shown here. In addition, the probe recognizes another sequence (left arrow) that was buried in the middle of chromosome 2 when chromosomes of ancestral apes fused to create the human chromosome. (To this day the great apes have one more chromosome pair than humans.) The point of fusion is very close to a rare fragile site that was observed in the chromosome 2 used in this study. The chromosome on the right is shown with the hybridization signal removed so that the fragile site, which appears as a gap (right arrow), can be discerned.*

From IJdo, J.W., Baldini, A., Wells, R.A., Ward, D.C., and Reeders, S.T. 1992. Genomics *12:833–835.*

Extracellular Factors Affecting Neuron Development



Louis F. Reichardt, Ph.D.—Investigator

Dr. Reichardt is also Professor of Physiology and of Biochemistry and Biophysics at the University of California, San Francisco. He received his Ph.D. degree in biochemistry from Stanford University for work on control of gene expression by a bacterial virus, the bacteriophage λ. Dr. Reichardt entered the field of neurobiology as a postdoctoral fellow in Paul Patterson's laboratory at Harvard University, where he studied factors that regulate the transmitter phenotype of individual neurons. Among his honors are a McKnight Scholars Award, a Sloan Award, and a Guggenheim Fellowship.

MY laboratory is investigating molecules in the extracellular environment of neurons (conducting nerve cells) that direct their development *in vivo*. These include trophic (nutritive) factors, exemplified by nerve growth factor (NGF), and molecules in the extracellular matrix or on the surface of cells that serve as substrates for the growth of axons (long nerve fibers). Such molecules help to regulate neuronal survival, axon growth, and synapse (nerve junction) formation.

Neurons require contact with targets to survive during development. Experimentally increasing or decreasing the volume of target tissue correspondingly increases or decreases neuronal survival. To explain these target influences, it has been proposed that target organs synthesize trophic factors required for the survival of the innervating neurons. Defects in the synthesis of these factors or in the neuron's ability to respond to them may explain some neurodegenerative disorders.

The key to understanding trophic factors is to identify and understand the actions of their receptors. One of the proteins that functions as a receptor for the family of NGF-related trophic factors is called the low-affinity NGF receptor. Antibodies to this protein have been prepared and used to show that it is distinct from a second protein that constitutes a distinct high-affinity receptor for NGF. Work in the laboratories of David Kaplan, Mariano Barbacid, and their colleagues has recently shown that this second protein is the ***trk*** proto-oncogene product, an NGF-tyrosine kinase. The same antibodies have been used to show that NGF-dependent survival and axon growth by response neurons do not require the low-affinity receptor. These antibodies are being used to examine the role of this class of receptor in mediating other responses to NGF.

Our laboratory has devoted considerable effort to identifying molecules that promote the growth of neuronal processes. We have tried to identify both the molecules that axons recognize in their environment and the receptors that neurons use for binding to these molecules. Our results have shown that laminin, an adhesive protein, is by far the most active of the glycoproteins that cells secrete into the extracellular matrix. Other proteins that have similar activities include fibronectin, thrombospondin, and vitronectin.

We have also shown that neurons utilize a family of receptors, the integrins, to bind to laminin and other matrix glycoproteins. Distinct receptors appear to mediate neuronal adhesion and growth cone motility on laminin, fibronectin, and collagen. In the past two years we have purified several of these proteins and isolated clones encoding subunits of the receptors. Specific antibodies to these subunits have been prepared and are being used to study their regulation. Of particular interest, we have identified the receptors used by both peripheral neurons and retinal neurons to interact with laminin, a heterotrimer assembled from three different gene products—A, B1, and B2. Josh Sanes and Eva Engvall have shown that at least two different A-like genes and two different B1-like genes exist and are differentially expressed, making it possible to assemble four different isoforms of laminin, which are differentially distributed in embryos. We have shown that individual integrin heterodimers distinguish between these isoforms, making it possible for cells to exhibit different responses to individual isoforms. Work in the past year has also identified candidate receptors for thrombospondin and tenascin.

Evidence suggests that the activity of integrin receptors may modulate the behavior of axons *in vivo*. In studies on regulation of $\alpha_6\beta_1$, the major laminin receptor in the neuroretina, we have shown that receptor function can be regulated on several levels. First, expression of the genes encoding the two subunits is regulated. Retinal ganglion cells lose responsiveness to laminin when they contact their synaptic partners in the optic tectum, and this reflects down-regulation of expression of the α_6 gene. Second, other neurons in the retina modulate responses to laminin by exhibiting changes in the activity of integrin receptors on their surfaces. These changes occur rapidly and can be modulated by external and internal agents. Yet a third level of regulation ap-

pears to modulate the signals conveyed by integrin receptor binding to laminin and other proteins. These signals can be modulated by trophic factors, such as NGF.

In addition to proteins secreted into the extracellular environment of cells, neurons also recognize and use for axon extension integral membrane proteins on the surfaces of different cell types. Recent work in our laboratory has identified some of the major proteins that neurons use for axonal growth in the central and peripheral nervous systems. Thus neurons have been shown to use both extracellular matrix and cell-cell adhesion molecules for growth of axons on Schwann cells, the major cell type with which they interact in peripheral nerves. A similar combination of adhesive interactions promotes growth of retinal axons on astroglia. The movements of growth cones of retinal ganglion cells have been directly monitored *in vivo* by time-lapse microscopy. Injection of antibodies to individual cell adhesion molecules has been shown to alter dramatically the behavior of these growth cones.

In studies on interactions of neurons with other cells in the brain, it became clear that additional adhesion molecules, not yet identified or characterized, must be important. We have used molecular biological methods to identify two of these with localized expression patterns in the brain. One of these, the integrin $\alpha_8\beta_1$, is localized on axons in many of the major axon tracts of the brain. Its concentration on axons in these tracts suggests that it mediates interactions between these axons. A potential ligand for this integrin that is also localized in these tracts has been identified. The second novel adhesion-promoting molecule is B-cadherin, a Ca^{2+}-dependent adhesion molecule that is also expressed in the nervous system. It appears to be concentrated in specific cells in several areas of the brain. Some but not all neurons are able to interact with it. Both molecules are also expressed in some areas outside the nervous system.

Future work will focus on determining how these neuronal receptors act to promote growth cone motility, how they are regulated during development, and what their importance is in regulating the growth of axons during development and regeneration.

Molecular Genetics of RNA Processing and Behavior

Michael Rosbash, Ph.D.—Investigator

Dr. Rosbash is also Professor of Biology at Brandeis University and Adjunct Professor of Molecular Biology at Massachusetts General Hospital, Boston. He received his Ph.D. degree in biophysics from the Massachusetts Institute of Technology and was a postdoctoral fellow at the University of Edinburgh, where he studied with J. O. Bishop. Dr. Rosbash was a Guggenheim Fellow in Paris, France.

MY laboratory is interested in two fundamental problems. Our earliest and foremost interest is the molecular genetics of RNA processing. For this subject our principal experimental system is the budding yeast *Saccharomyces cerevisiae*, which is amenable to both genetic and biochemical attack. Our more recent interest is the molecular genetics of behavior—in particular, circadian rhythms. This problem is addressed in the fruit fly *Drosophila melanogaster*, because the organism is amenable to behavioral as well as biochemical and genetic approaches.

Within the area of RNA processing, we are most interested in understanding certain aspects of pre-messenger RNA splicing, the process by which the undesirable sections of a pre-mRNA molecule are removed and the remaining "sense" sections sewn together. Our interests are primarily focused on the more biological aspects of the problem. These include how the places in the molecule to be cut—the two splice sites in the case of a pre-mRNA with a single intron—are defined. They also include how splice site partners, in the case of a pre-mRNA with multiple introns, are specified.

The latter question is particularly puzzling, because most of these splice sites appear similar. Yet there is clearly an order to the process, although the basis for this order is not apparent. The adjective "biological" is used to distinguish these issues from more "chemical" considerations, such as how the active site of the splicing enzyme is formed and how the efficiency and specificity of the splicing reaction are dictated.

We are addressing these issues of splice site definition and partner assignment by examining the interactions between a pre-mRNA substrate and splicing factors. Although some of the work is done *in vivo*—that is, in intact cells where the interactions are inferred from their consequences—most of our efforts have concentrated on interactions that take place during *in vitro* pre-mRNA splicing in a whole-cell yeast extract. We have focused particularly on the earliest interactions, those that apparently reflect recognition of the pre-mRNA substrate by the splicing machinery.

Our studies indicate that the factor U1 small nuclear ribonucleoprotein (snRNP) plays a prominent role in these early interactions. Consequently we have expended considerable effort in characterizing this snRNP and its constituents, as well as the pre-mRNA–U1 snRNP interaction. Surprisingly, both ends of the intron interact with U1 snRNP, suggesting that certain aspects of splice site recognition, if not splice site partner assignment, are already defined early during the spliceosome assembly process, well before the cleavage and ligation steps of the actual splicing process take place. This work is supported by a grant from the National Institutes of Health.

We are also interested in another biological aspect of RNA processing—namely, how mRNA is transported from the nucleus, where it is synthesized, to the cytoplasm, where it is translated into protein. This transport problem interfaces with the splicing process, since RNA needs to be transported to the cytoplasm but usually not before the splicing is completed. Otherwise, incompletely spliced molecules would be prematurely transported, which would give rise to untranslatable pre-mRNAs in the cytoplasm.

The problem of RNA transport is poorly understood, and even less well understood in yeast than in mammalian cells. In yeast, however, there is the possibility of addressing the problem with genetic tools. At present we are localizing pre-mRNA and splicing factors within the yeast nucleus, in an attempt to define a cytological path that the RNA may follow in leaving the nucleus. The goal is to use existing temperature-sensitive mutants and to uncover new ones, both to study RNA transport and to define some of the gene products important for this process.

Rhythms

Our goals in this project are to define the biochemical machinery that underlies the mysterious yet ubiquitous process of circadian rhythmicity. We are using genetics and biochemistry to define candidate genes and gene products that may participate in fundamental aspects of these rhythms. Our entrée into the process is the *period* gene (*per*) of *Drosophila melanogaster*.

Mutants in this gene, originally identified more than 20 years ago, have profound effects on circadian rhythms of locomotor activity and of eclosion (emergence of adults from the pupal case).

Recently we discovered that the *per* gene products, mRNA as well as protein, undergo circadian fluctuations in level during the circadian cycle. These observations and others indicated that there is a feedback loop in which the *per* protein negatively affects the level of its own mRNA. Since the *per* mRNA also gives rise to the *per* protein during translation, this putative feedback loop contains all of the elements required to define a circadian clock, including a substantial (and mysterious) delay between the mRNA and protein accumulation curves. We are studying several aspects of this loop in an attempt to confirm (or refute) its importance to the circadian clock.

Because no close relatives of the *per* protein with a known biochemical function are found in the database, we cannot say for certain what gene family it belongs to or what biochemical function it serves. We suspect, however, that it is a transcription factor, or serves to modulate transcription, and that its effect on its own mRNA levels is quite direct. A current goal is to test this hypothesis and to define the biochemical function of the *per* gene product, especially insofar as rhythms are concerned.

With the support of a grant from the National Institutes of Health, we are also in the process of defining and studying several other genes that affect rhythms. Although these genes are in a less well developed state of examination than the *per* gene, some appear interesting and may provide additional insight into clock mechanisms.

Molecular Mechanisms of Transcription, Regulation, and Development of the Neuroendocrine System

Michael G. Rosenfeld, M.D.—Investigator

Dr. Rosenfeld is also Professor of Medicine in the Eukaryotic Regulatory Biology Program at the University of California, San Diego, School of Medicine. He received his undergraduate degree from the Johns Hopkins University and his medical degree from the University of Rochester. His internship and medical residency were completed at Barnes Hospital, St. Louis. Before accepting his current position, he received postdoctoral training at NIH. Dr. Rosenfeld also holds an adjunct position at the Salk Institute.

OVER the past year our central research focus has been the determination of molecular mechanisms that induce specific neuroendocrine phenotypes and the further definition of signal transduction pathways that lead to regulated patterns of gene expression.

The neuroendocrine system coordinates the complex pattern of regulation necessary to achieve the precise temporal, spatial, and homeostatic patterns of gene expression required by complex organisms. Development of the central nervous system and endocrine organs involves precise patterns of responses to morphogens and other regulatory signals that ultimately establish the intricate patterns of neural and endocrine phenotypes. The cloning and analysis of genes encoding receptors and cell-specific transcription factors have permitted an initial definition of developmental and regulatory strategies.

The anterior pituitary gland has provided a suitable model to investigate the molecular basis for generating specific cell phenotypes in an organ. The rat genes for growth hormone and prolactin (hormone that stimulates and sustains lactation) exhibit precisely restricted expression in the cells of origin, somatotrophs and lactotrophs, respectively. We found that prolactin gene expression is dictated by a distal enhancer and a proximal region, each containing at least four critical cell-specific elements. These two domains, each capable of targeting tissue-specific gene expression, act synergistically to generate high levels of prolactin gene expression in transgenic mice. Similarly, growth hormone gene expression involves the action of related cell-specific cis-active elements. A 33-kDa cell-specific transcription factor, referred to as Pit-1, was characterized and its encoding cDNA defined.

Bacterially expressed Pit-1 specifically and with high affinity binds to prolactin and growth hormone promoters. Additional elements and factors are required to achieve the full physiological levels and restricted patterns of expression of the prolactin and growth hormone genes. For example, the estrogen receptor synergistically acts with Pit-1 in activation of the rat prolactin gene's distal enhancer.

Pit-1 is a member of a family of regulators that contain a POU domain consisting of a variant homeodomain and a second, 76–amino acid sequence of striking homology. We have found that the 76–amino acid POU-specific domain functions in high-affinity DNA binding, in conferring site-specificity, and in protein-protein interactions critical for transcriptional activation by Pit-1. The major transcription-activating domain of Pit-1 is a 70–amino acid, N′-terminal, serine, threonine-rich sequence, distinct from recognized motifs. A detailed structure-function analysis has suggested that the determinants of DNA binding by the Pit-1 POU-domain protein are distinct from those of the classical homeodomain proteins, and that Pit-1 DNA binding is regulated by phosphorylation of a specific, conserved residue in the POU homeodomain.

A genetic approach was utilized to determine the functional role of Pit-1 during organogenesis. Based on the demonstration of disrupted patterns of the Pit-1 gene in genetic dwarf mice, it has been established that this POU-domain protein acts as a developmental regulator to determine patterns of commitment, progression, and proliferation of three specific cell types in the anterior pituitary gland. In the case of lactotrophs and somatotrophs, proliferation is restricted in the Pit-1–defective animal cell type, while the thyrotroph provides an initial example of an established cell phenotype that disappears at the time the initial Pit-1 protein is expressed. These data indicate that one role of a developmental transcriptional regulator can be survival of an established cell type.

A strategy was devised to isolate new members of the POU-domain gene family. Ten new members were identified in neural tissues. Two, referred to as *Brn-1* and *Brn-2*, exhibit virtually identical patterns of expression in the central nervous system, though *Brn-1* is clearly expressed in the medullary zone of the kidney while *Brn-2* is not. A third member, *Brn-3,* is predominantly expressed in the peripheral nervous system. *Tst-1* transcripts are present in mammalian brain cells and in testis. Subsequently other

POU domains have been identified, each expressed in a unique pattern early in development and during organogenesis.

Most POU-domain genes were widely expressed in all levels of the neural tube (including the retina) during early development, and hybridization in the ventricular (proliferative) zone of the neuroepithelium is evident for all four transcripts at all levels. The time course of anatomic restriction in the developing neural tube is distinct, and the patterns for each gene product tend to reflect the adult loci of expression. In addition, reactive transcripts for some of these genes are also expressed during mammalian neurogenesis.

We have recently identified many additional family members in the brain and have obtained initial evidence that at least some members of the family can bind to specific elements in distinct classes of neuronally expressed genes. We find that neurally expressed POU-domain proteins can be considered to represent distinct related families, each of which binds to related, but distinct, DNA response elements. Putative target genes have been identified for several of these factors, and analysis of their function has revealed both inhibitory and stimulatory domains.

Interactions between diverse transcription factors generate heterodimers that exert distinct patterns of gene activation. Thus, in the case of the retinoic acid receptor, we find that a series of cell type–specific coregulators impose novel hierarchies of binding site preferences. Unique positive and negative transcriptional regulators impose variable patterns of gene activation, potentially contributing to the refinement of phenotypic variance required in the central nervous system. One class of coregulators has been characterized by expression cloning techniques.

Development of the neuroendocrine system is also initially regulated by post-transcriptional strategies. We have provided evidence that a gene of the neuroendocrine system, the calcitonin/CGRP (calcitonin gene-related peptide) gene, contains genomic regions that represent discrete hormone-encoding domains. The ultimate expression of these domains is dependent on alternative RNA-processing events that differentially include or exclude specific exons encoding certain components of polypeptide regulators in the mature mRNA products. The rat and human calcitonin/CGRP genes contain six exons. More than 95 percent of the mature transcripts in thyroid C cells, encoding calcitonin, are produced by splicing of the first three to four exons. CGRP mRNA is the only detectable mature transcript in rat neuronal tissue, which appears to reflect the actions of a specific regulatory machinery controlling post-transcriptional RNA splice acceptor choice. CGRP appears to be an important regulator of blood pressure.

Current investigations continue to explore these novel aspects of transcriptional and post-transcriptional regulatory strategies in neuroendocrine gene expression. Results may provide new ways of studying the problem of organogenesis.

Development of the Drosophila *Visual System*

Gerald M. Rubin, Ph.D.—Investigator

Dr. Rubin is also John D. MacArthur Professor of Genetics at the University of California, Berkeley, and Adjunct Professor of Biochemistry and Biophysics at the University of California, San Francisco, School of Medicine. He received his B.S. degree in biology from the Massachusetts Institute of Technology and his Ph.D. degree in molecular biology from the University of Cambridge. Dr. Rubin's postdoctoral work was done at Stanford University with David Hogness. He has held faculty positions at Harvard Medical School and the Carnegie Institution of Washington. Dr. Rubin is a member of the National Academy of Sciences and counts among his other honors the American Chemical Society Eli Lilly Award in Biological Chemistry.

RESEARCH in our laboratory is directed toward studies of differentiation and gene regulation in the developing nervous system. Our experimental approach involves studying genes whose mutations disrupt neural development. During the past year, we have focused our work on several genes important for the determination of cell fates in the developing retina of the fruit fly *Drosophila*.

Two very different but not exclusive mechanisms can account for the selection of distinct developmental pathways. First, cells may be programmed in a lineage-dependent manner by the asymmetric partitioning of determinants during cell division. Different developmental pathways are then selected in the daughter cells in response to the different localized determinants. Alternatively, cellular differentiation may occur in a lineage-independent manner, where the position that a cell occupies in a developing field determines its fate. In this case, diffusible substances, such as hormones, or interactions with adjacent cells are the primary determinants of cellular differentiation. Although the mechanisms used to read and interpret such positional information are largely unknown, short-range cellular interactions are thought to be of principal importance in a wide variety of developmental phenomena.

The compound eye of *Drosophila melanogaster* is an attractive system to study the mechanisms underlying lineage-independent developmental decisions, since it consists of a small number of different cell types that develop in a lineage-independent manner. The compound eye is a two-dimensional array of 800 repeating units, or ommatidia. Each ommatidium contains 8 photoreceptor cells as well as 12 nonneuronal accessory cells. Each photoreceptor cell has a distinct cellular identity, based on both its position within the ommatidium and its projection pattern to the optic lobes of the brain. The stereotyped arrangement of this small number of nerve cells, together with the dispensability of the visual system under laboratory conditions, makes the compound eye an attractive model system to study genes involved in the specification of nerve cells.

Assembly of ommatidia begins in an initially unpatterned monolayer of epithelial cells, the eye imaginal disc. Ommatidial assembly does not occur synchronously throughout the disc but instead begins at the posterior edge and progresses anteriorly. Eye discs removed from larvae just prior to pupariation show a smoothly graded series of ommatidia at different stages of development, covering just over half of the disc. Examination of individual cells in the forming ommatidia has shown that the photoreceptors differentiate in a fixed sequence, beginning with the central R8 photoreceptor and proceeding pairwise with R2 and R5, R3 and R4, R1 and R6, and finally R7.

The fate of a cell within a developing ommatidium appears to be governed by the specific combination of signals received by that cell from its immediate neighbors. We would like to understand how such signals are generated, received, and interpreted. Our approach has been to study mutations that specifically disrupt these processes, as illustrated by our studies of the *sevenless* gene.

The *sevenless* gene is essential for the development of a single type of photoreceptor cell. In the absence of proper *sevenless* function, the cells that would normally become the R7 photoreceptors become instead nonneuronal cells. Previous morphological and genetic analysis has indicated that the product of the *sevenless* gene is involved in reading or interpreting the positional information that specifies this particular developmental pathway. We have isolated and characterized the *sevenless* gene. Our data indicate that *sevenless* encodes a transmembrane protein with a tyrosine kinase domain. The structural analogies between the *sevenless* protein and certain hormone receptors suggest that developmental pathway selections dependent on cell-cell interactions may involve molecular mechanisms similar to physiological or developmental changes induced by long-range diffusible factors.

To investigate the role of the *sevenless* protein

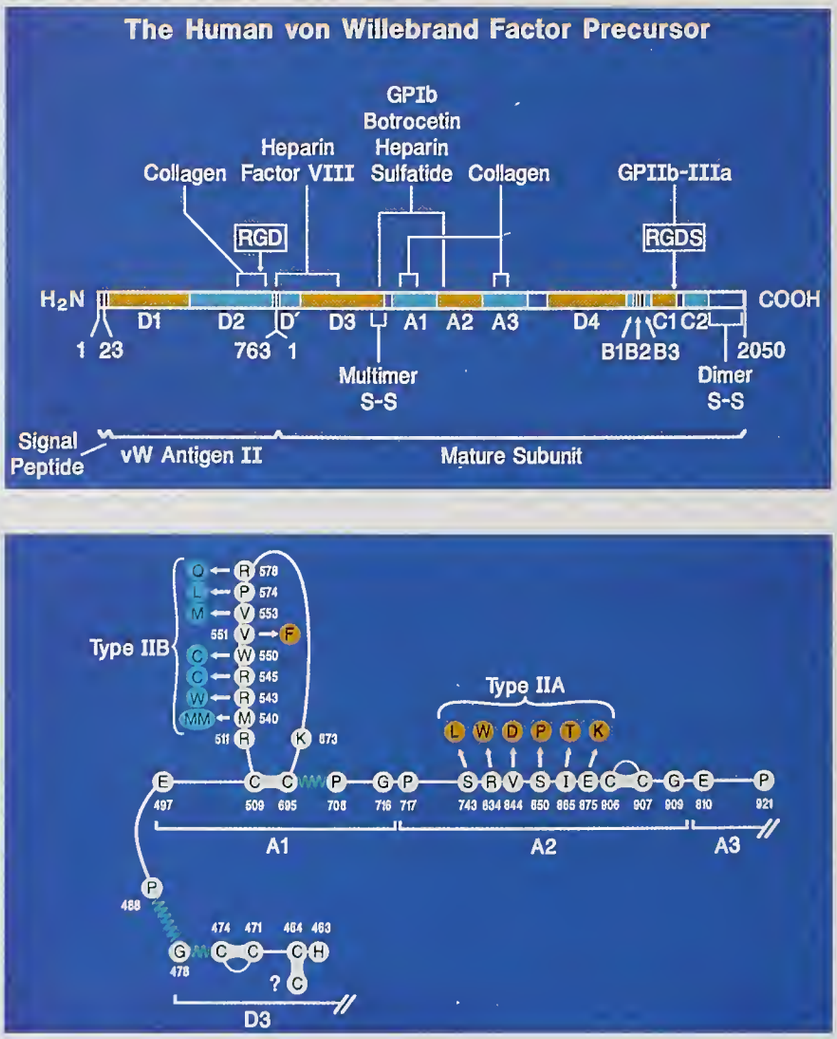
in R7 development, we have examined the pattern of *sevenless* expression in the developing retina and have studied the effects of experimentally altering this pattern. By transiently expressing the *sevenless* protein, we have shown that there is only a brief period during eye development when the *sevenless* protein is required for the formation of the R7 photoreceptor. Our results are consistent with the proposal that *sevenless* directly reads positional information required to specify the R7 developmental pathway.

A major current goal is to elucidate the intracellular signal transduction pathway that is activated by stimulation of the *sevenless* protein. That is, how does activation of the *sevenless* tyrosine kinase instruct a cell to become an R7 photoreceptor? Defining the biologically relevant substrates of tyrosine kinase receptors has been a long-standing and difficult problem. We have taken a genetic approach toward identifying genes whose products act downstream of *sevenless*, including those that might be direct substrates for the *sevenless* kinase.

We have utilized two strategies. First, we have looked for other mutations that give the same phenotype as *sevenless*—transformation of the R7 cell into a nonneuronal cell type. In this way we isolated the *seven-in-absentia* (*sina*) gene. Function of the *sina* gene is required only in R7 for correct R7 cell development. The *sina* protein, which has a potential metal-binding domain, is localized in the nuclei of several ommatidial precursor cells, including R7, and *sina* expression in R7 appears before R7 overtly begins to differentiate. These data indicate that the *sina* gene product is necessary at a stage in the determination of R7 cell fate when R7 receives and interprets developmental signals from neighboring cells and possibly acts by regulating gene expression.

Second, we utilized a crippled *sevenless* protein, whose activity is just barely adequate to specify R7 cell development, to establish a highly sensitive assay for other components of this signal transduction pathway. Using this assay we looked for other genes in which a 50 percent reduction of the level of their protein products resulted in a failure to specify R7 cells. In this way we have identified seven genes that appear to encode products that act to interpret the signal mediated by the *sevenless* receptor.

The putative products of two of these seven genes have been identified. One encodes a *ras* protein. The *ras* oncogene is implicated in as many as 30 percent of human tumors. The *ras* proteins exist in two different states: an inactive GDP-bound state and an active GTP-bound state. The active *ras* protein transmits a signal by interaction with unidentified cellular targets. The other locus whose product we have identified encodes a protein that is homologous to the *Saccharomyces cerevisiae CDC25* protein, an activator of GDP-GTP exchange by *ras* proteins. These results suggest that the stimulation of *ras* protein activity is a key element in the signaling by *sevenless* and that this stimulation may be achieved by activating the exchange of GTP for bound GDP by the *ras* protein. The evolutionary conservation of the *ras* signaling pathway suggests that studies in *Drosophila* could provide clues to the role of *ras* in oncogenesis and developmental abnormalities in humans.



Top: Structure-function relationships of von Willebrand factor. The factor precursor comprises four types of homologous domains (A–D), repeated two to five times each. The mature subunit consists of 2,050 amino acids. Binding sites are indicated for collagen, heparin, factor VIII, platelet glycoprotein Ib, botrocetin, sulfatide, and platelet glycoprotein IIb–IIIa. The locations of two Arg-Gly-Asp–containing sequences, RGD and RGDS, are indicated.

Bottom: Mutations, indicated by brackets, in exon 28 of the gene encoding von Willebrand factor cause von Willebrand disease types IIA and IIB. The segment of the factor shown includes amino acid residues 463–921, designated by single-letter code. The positions of repeated domains D3 and A1–3 are indicated. The green zigzag segments from Cys474 to Pro488 and Cys695 to Pro708 indicate regions proposed to interact directly with platelet glycoprotein Ib. One proposed type IIA mutation (Val551 → Phe, orange circle) occurs in the region of the type IIB mutations.

From Sadler, J.E. 1991. J Biol Chem *266:22777–22780.*

The Regulation of Blood Coagulation

J. Evan Sadler, M.D., Ph.D.—Associate Investigator

Dr. Sadler is also Associate Professor of Medicine and of Biochemistry and Molecular Biophysics at Washington University School of Medicine, St. Louis. He obtained his undergraduate degree in chemistry from Princeton University. He then attended Duke University, where he received first his Ph.D. degree in biochemistry with Robert Hill and then his M.D. degree. Following his internship and residency in medicine at Duke University Medical Center, Dr. Sadler was a Hematology Fellow in the laboratory of Earl Davie at the University of Washington, Seattle.

UNDER normal circumstances, blood clots occur only at sites of vascular injury, and unnecessary clots are dissolved promptly. Inappropriate blood clots can cause devastating illness, such as stroke and myocardial infarction. Abnormal thrombosis also complicates many common diseases, including certain cancers and infections.

In the blood, proteins and small cells called platelets are required for clot formation. The endothelial cells that line all blood vessels and circulating white blood cells are not, however, passive bystanders in these reactions, but actively promote or inhibit clotting. Compounds that are produced during inflammation modulate these cellular activities.

We are investigating the structure, function, regulation, and evolution of proteins that control blood coagulation. Our goal is to understand how these opposing tendencies—to stimulate or to inhibit clotting—are balanced to achieve normal hemostasis and prevent dangerous thrombosis. These studies will increase our knowledge of the interaction between blood coagulation and inflammation and may provide a foundation for the design of new therapies for thrombotic disorders. Studies of thrombomodulin, thrombin, and tissue factor are supported by a grant from the National Institutes of Health.

von Willebrand Factor and von Willebrand Disease

The von Willebrand factor (vWF) is a blood protein that is made by endothelial cells and is required for normal platelet function. vWF also binds to and stabilizes blood coagulation factor VIII, the factor that is deficient in classical hemophilia. The structure of vWF was determined indirectly by cDNA cloning: vWF contains 12 repeated domains that belong to four families of ancestral sequences. Hereditary deficiency of vWF, or von Willebrand disease, is the most common genetic bleeding disorder of humans. Mild abnormalities of vWF function can be detected in nearly 1 percent of the population.

We determined the structure of the human vWF gene and also of a related pseudogene that has diverged recently from the authentic vWF gene. This allowed us to investigate von Willebrand disease at the level of gene sequence. We characterized deletions of the vWF gene that cause severe von Willebrand disease in five unrelated patients. These particular patients treat transfused vWF as a foreign protein and make inhibitory antibodies to it. Deletions in the vWF gene predispose to the formation of such antibodies.

These studies were extended to include patients with variants of von Willebrand disease who make a defective vWF molecule. Among more than 30 unrelated patients, 11 different mutations were characterized within a single exon of the vWF gene. Six of these mutations are within a small vWF domain that appears to modulate the affinity of vWF for platelets; of these six mutations, five cause a paradoxical increase in binding and the sixth causes a decrease in binding. The remaining five mutations are within an adjacent domain of the protein and cause loss of function, either by impairing vWF biosynthesis or increasing vWF degradation. The association of severe bleeding with both increased and decreased function illustrates the importance of balanced vWF function for normal hemostasis.

A recently described variant of von Willebrand disease, recognized in a patient from Normandy, France, is characterized by defective binding of vWF to blood coagulation factor VIII. In such patients factor VIII is unstable, and this results in a secondary factor VIII deficiency that mimics classical hemophilia. Among several unrelated affected families, three different mutations were identified in the factor VIII–binding site of vWF. The corresponding recombinant vWF proteins exhibited the same defect in factor VIII binding as natural vWF Normandy, confirming that these mutations cause the disease. The genetic defects of these and other such patients provide insight into structure-function relationships of vWF and may suggest new therapeutic strategies to inhibit or augment vWF function.

Thrombomodulin and Thrombin

Thrombomodulin is a protein of the endothelial cell surface that binds thrombin, a blood clotting enzyme. Because of its effects on thrombin activity, thrombomodulin is an essential natural anticoagulant. Several inflammatory mediators decrease the expression of thrombomodulin by endothelial cells. Understanding this process may help us understand the abnormal blood coagulation that accompanies much human disease. We have cloned and expressed variants of recombinant human thrombomodulin in a variety of cultured cell lines. These cells have been used to define the structural requirements for thrombomodulin cofactor activity.

Thrombin is a protease enzyme that is required to form blood clots from fibrinogen. Thrombin also can inhibit clotting by first binding to thrombomodulin and then digesting certain other blood coagulation factors. Disruption of thrombin's normal balance between promoting and inhibiting clotting can, in principle, cause either bleeding or thrombosis. We have constructed mutant forms of thrombin with predominantly anticoagulant or procoagulant activities. Such mutant thrombins help to define the structural basis for the different activities of thrombin. In addition, they provide reagents to test the physiological importance of specific thrombin activities, and mutant thrombins may also be found to have therapeutic procoagulant or anticoagulant properties.

Tissue Factor

Tissue factor, a cell surface protein that is found on many cells that do not normally contact the blood, is the most important physiological initiator of blood coagulation. When blood vessels are damaged, tissue factor is exposed to and binds to blood coagulation factor VII. The factor VII–tissue factor complex then initiates a cascade of reactions that cause blood to clot. Both monocytes and endothelial cells express tissue factor activity in response to many stimuli, and this contributes to the abnormal thrombosis that accompanies systemic infections.

We have isolated cDNA clones for human tissue factor and localized the gene to chromosome 1. These clones have been used to study the regulation of tissue factor in endothelial cells. Tumor necrosis factor, a protein made during inflammation, causes a dramatic but transient increase in tissue factor activity. This appears to be the result of activating the gene, which is normally silent in endothelium. A second level of control can be employed to amplify this response. Tissue factor mRNA normally is degraded very rapidly, but it is stabilized by some agents that stimulate endothelial cells, and this may contribute to the induction of tissue factor activity during inflammation. We are currently studying the structural basis for this regulation of tissue factor mRNA degradation.

The systems we are investigating provide abundant opportunities to answer biological questions concerning the regulation of blood coagulation and to approach fundamental questions related to signal transduction, gene expression, and protein structure-function relationships. These studies may illustrate how several proteins can be coordinately regulated to promote blood clotting reactions on the vascular endothelium during inflammation. We will continue to explore the mechanisms by which vWF, thrombomodulin, thrombin, and tissue factor are regulatcd and will extend this work to other endothelial cell proteins that can promote or inhibit thrombosis.

Molecular Mechanism of Transmembrane Signal Transduction by G Protein–coupled Receptors

Thomas P. Sakmar, M.D.—Assistant Investigator

Dr. Sakmar is also Assistant Professor at the Rockefeller University. He received his A.B. degree in chemistry and his M.D. degree from the University of Chicago. He completed a medical residency at Massachusetts General Hospital, Boston, and conducted postdoctoral research in the laboratory of H. Gobind Khorana at the Massachusetts Institute of Technology.

IN our laboratory the vertebrate visual proteins rhodopsin and transducin serve as a model system for structure-function studies on the molecular mechanism of transmembrane signaling. These visual proteins are members of a superfamily of related G proteins (guanine-binding regulatory proteins) and G protein–coupled receptors. Light-activated rhodopsin catalyzes guanine nucleotide exchange by transducin, which ultimately leads to a change in membrane cation conductance and a neural signal. Our approach is to reconstitute heterologously expressed rhodopsin and transducin in defined *in vitro* systems and to use biochemical and biophysical methods to probe site-directed mutants.

Our current interests include structure-function relationships in rhodopsin. For example, we are studying the ground state structure of the receptor, the interactions between specific amino acid residues and the 11-*cis*-retinal chromophore that control spectral properties and photochemistry, the mechanism by which a photochemical signal is transmitted from the core of the receptor to the surface, and the specific domains on the cytoplasmic surface that bind and activate transducin.

We employ a multifaceted approach, including the use of a variety of complementary spectroscopic techniques. For example, the structure and environment of the retinal chromophore in rhodopsin and its photointermediates can be studied with resonance Raman spectroscopy. In a continuation of collaborative work with Steven Lin and Richard Mathies, we obtained microprobe resonance Raman spectra of solutions containing only microgram quantities of mutant pigments.

The results confirmed and supplemented our earlier observations concerning the role of a specific carboxylate group in rhodopsin that acts to stabilize the positive charge of the protonated Schiff base chromophore linkage. A model of the chromophore binding pocket of rhodopsin was proposed and is being used to direct further studies into the mechanism of wavelength regulation by visual pigments.

We are also interested in identifying specific domains of rhodopsin and transducin involved in binding and activation. Flash photolysis studies of site-directed rhodopsin mutants had previously shown that at least the second and third cytoplasmic loops of rhodopsin are involved in activation of bound transducin. Some cytoplasmic mutations prevent transducin binding as well. Recently we have developed a spectrofluorimetric method designed to allow simultaneous illumination and measurements of rhodopsin-transducin interactions by intrinsic fluorescence.

Rhodopsin-catalyzed binding of GTP, or a GTP analogue to transducin, results in a large increase in its intrinsic fluorescence. Mixtures of transducin and rhodopsin can be assayed by this method to determine the kinetic rate constants of their interaction and to evaluate the specific effects of mutations. A series of site-directed mutants of rhodopsin with alterations in their cytoplasmic domains have been studied. The results may be relevant to other seven-transmembrane helix receptors that couple to G proteins and play roles in cellular physiology, growth, development, and differentiation in the nervous system.

Molecular Genetics of Development in Drosophila

Shigeru Sakonju, Ph.D.—Assistant Investigator

Dr. Sakonju is also Assistant Professor of Human Genetics at the University of Utah School of Medicine. He received a B.A. degree from Columbia Union College and a Ph.D. degree in biology from the Johns Hopkins University, having developed his doctoral thesis in the Department of Embryology at the Carnegie Institution of Washington, Baltimore, with Donald Brown. He was a Helen Hay Whitney Postdoctoral Fellow with E. B. Lewis at the California Institute of Technology and at Stanford University with David Hogness.

DURING the development of organisms, the fertilized egg undergoes many divisions to produce a multicellular body. In the fruit fly *Drosophila melanogaster*, the body is made up of several fused segments in the head, 3 thoracic segments with wings and legs, and 10 abdominal segments, each showing unique characteristics. This basic pattern of body segments, invariant from generation to generation, is dictated by a genetic blueprint within the organism's own genome.

The characteristics, or identity, of each body segment are determined by the activities of the so-called homeotic genes. When these do not function properly, a body segment or group of segments transforms to take on the characteristics of another segment. Thus homeotic genes can be thought of as master regulatory switches that trigger the genetic circuits necessary to form normal body patterns. Genes similar to the homeotic genes of the fruit fly are found in other organisms, including humans. The goal in my laboratory is to understand how the homeotic genes accomplish their task at the molecular level.

In *Drosophila*, three homeotic genes—*Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*)—are responsible for determining the characteristics of two thoracic and nine abdominal segments. These genes are located in a chromosomal region called the bithorax complex (BX-C). Flies carrying mutations in any one of the three BX-C genes show characteristic transformations of body segments. By observing which segments are transformed in these mutants, we know that *Ubx* is required in two thoracic and eight abdominal segments, *abd-A* in the second through eighth abdominal segments, and *Abd-B* in the fifth through ninth abdominal segments.

To learn what homeotic proteins do in the cell, we have focused on the proteins encoded by *Ubx* and *abd-A*. Many homeotic proteins, including these, have been shown to bind DNA *in vitro*, suggesting that they act by binding to transcriptional signals of other genes to regulate expression. Whether this is how they act in living organisms has not been directly shown. Therefore we have chosen to study the regulation of a potential target gene of Ubx and abd-A proteins, called *Antennapedia* (*Antp*), as a paradigm for the mechanism of homeotic protein actions.

We have shown that both Ubx and abd-A proteins bind to a number of sites on the DNA segment that contains the signals necessary for transcription of the *Antp* gene. We have further shown, by creating mutations in the binding sites and assaying their effects, that this binding is essential for repressing the *Antp* gene expression in the embryo. Our study therefore provides direct evidence that homeotic proteins do turn on or off transcription of their target genes.

What are the target, or downstream, genes regulated by homeotic proteins? Little is known about potential target genes, even though many are thought to exist. To identify such targets, we have utilized a method, called the enhancer trap, of detecting genes with expression patterns that suggest homeotic gene regulation. A number of candidates have been isolated, and we are testing them for further evidence of being the downstream genes. We have shown that the expression of several candidate genes is abnormal in the mutant embryos that do not produce Ubx protein, suggesting that the protein does regulate this gene.

We are also interested in determining how the homeotic genes with a relatively small number of encoded proteins can specify many unique body segments. One answer lies in the fact that these genes are expressed in different but overlapping sets of segments. For example, of the three homeotic proteins from the BX-C, only Ubx protein is detected in the second and third thoracic segments; both Ubx and abd-A proteins are present in the second through fourth abdominal segments; and all three BX-C proteins are found in the fifth through eighth abdominal segments.

These three combinations would of course define only three segment identities if all cells within segments were expressing the same combination of the homeotic genes. In fact, cells within a segment do not express the same combination.

We find that some cells within a given abdominal segment express abd-A protein exclusively or predominantly; other cells, Ubx protein; and still other cells, Abd-B protein. This kind of "mosaic" expression can, in theory, specify an unlimited number of segmental identities. We have therefore asked the following question: Is the mosaic expression necessary for correct specification of these identities?

To answer this question, we altered flies genetically so that homeotic proteins could be expressed at will. When Abd-B protein is expressed in all cells, thoracic and abdominal segments exhibit some characteristics of the eighth abdominal segment. When both Ubx and Abd-B proteins are expressed simultaneously, thoracic segments transform to the first abdominal segment, which is specified by *Ubx*, but abdominal segments remain largely unaltered. Therefore, surprisingly, we do not observe completely nonsensical identities, as might have been expected from this abnormal situation, suggesting that mosaic expression is not required, at least at a gross level, for correct specification of segments.

Generating a Repertoire of Antigen-Specific Receptors

David G. Schatz, Ph.D.—Assistant Investigator

Dr. Schatz is also Assistant Professor of Immunobiology at Yale University School of Medicine. He received undergraduate degrees in molecular biophysics and biochemistry from Yale University and in philosophy and politics from Oxford University. His Ph.D. degree and postdoctoral training were done with David Baltimore at the Massachusetts Institute of Technology and the Whitehead Institute.

CELLS of the immune system act in concert to protect against infectious agents and transformed (malignant) cells. At the heart of this protective system are the antigen receptor molecules found on B and T lymphocytes: the immunoglobulin (Ig) and the T cell receptor (TCR). Each lymphocyte expresses a distinct receptor molecule that confers on the cell a unique antigen specificity. The millions of different genes needed to encode these receptors are assembled from component gene segments by a site-specific process known as V(D)J recombination—so named for the V (variable), D (diversity), and J (joining) gene segments used in the reaction. V(D)J recombination is critical for the development of B and T lymphocytes and is the only site-specific recombination process known to occur in vertebrates. We are interested in two fundamental questions concerning such recombination: What is the biochemical mechanism of the reaction, and what molecular mechanisms regulate the reaction during lymphoid development?

This recombination reaction has been intensively studied since its discovery in 1976, yet little was known by the late 1980s about the enzymatic machinery (recombinase) that carried it out. Particularly frustrating was the inability to identify the gene or genes encoding the V(D)J recombinase, despite a detailed understanding of the substrates and products. While working with David Baltimore, I developed a novel genetic approach to the identification of these genes. Using a highly sensitive assay for V(D)J recombinase activity, I was able to demonstrate that gene transfer (transfection of genomic DNA) could activate the V(D)J recombinase in nonlymphoid cells, a surprising result implying that a single genetic locus was sufficient for the activation.

Marjorie Oettinger and I then isolated this genetic locus and quickly encountered a second surprise. The locus contained not one but two genes, which functioned together to activate the V(D)J recombinase. Indeed, a mixture of these recombination-activating genes, called RAG-1 and RAG-2, was thousands of times more potent than was either gene alone. We then went on to demonstrate that the two genes are only coexpressed in developing lymphocytes—in exactly those cells that are assembling Ig and TCR genes. This and a variety of other data suggest that RAG-1 and RAG-2 encode the critical, lymphoid-specific components of the V(D)J recombinase. Interestingly, we found that RAG-1, but apparently not RAG-2, is transcribed in neurons in the central nervous system, raising the possibility that RAG-1 has important roles in processes other than classic V(D)J recombination.

A central goal in our current research is to understand the enzymatic mechanism of V(D)J recombination, an elusive goal thus far, largely because efforts to reconstitute the reaction in a cell-free system have been unsuccessful. As a first step toward this goal, we are studying the biochemical and enzymatic properties of the RAG-1 and RAG-2 proteins. Our initial focus is on developing the necessary reagents—in particular, highly purified preparations of the proteins and antibodies that specifically recognize them.

We will use these tools to examine the biochemical properties of the RAG proteins, asking whether they exhibit the activities expected of proteins involved in recombination (e.g., topoisomerase, endonuclease, exonuclease, or ligase activities). We are also interested in determining if the RAG proteins bind to DNA, either nonspecifically or by interacting with elements of the V(D)J "recombination signal sequence" [the DNA element required for V(D)J recombination]. In addition, we hope to ascertain whether the RAG-1 and RAG-2 proteins interact with one another or with other proteins and how these interactions vary during lymphoid development and among different cell types.

Our second major focus is to understand the molecular mechanisms that regulate V(D)J recombination during lymphoid development. The assembly of Ig and TCR genes is a complex and highly ordered process regulated in part at the level of expression of the V(D)J recombinase. Using RAG-1 and RAG-2 as indicators, we are studying when and how the recombinase is turned on and then off again as B and T cells develop. Experiments performed in collaboration with Craig

Thompson (HHMI, University of Michigan) and Larry Turka suggest that signals transduced through membrane-bound TCR molecules may play a role in mediating the termination of V(D)J recombinase expression during T cell development.

V(D)J recombination is also regulated at the level of the availability, or accessibility, of the gene segments to the V(D)J recombinase. Accessibility in turn appears to be mediated by, or paralleled by, transcription of the unrearranged ("germline") gene segments. Analysis of germline transcription during B cell development has yielded important insights into the relationship between transcription and accessibility but has left unaddressed important questions concerning T cell development. We are interested both in how a stem cell (a cell with the potential to develop into multiple cell types) becomes committed to the T cell lineage and how the different sublineages of T cells are established. Since the regulation of V(D)J recombination is interwoven with these developmental decisions, we are examining the structure and developmental profile of germline TCR gene transcripts. We hope to determine the sequence of molecular events that lead to the assembly of TCR genes and to ask whether defects in this process contribute to the pathogenesis of immunological disease, particularly autoimmunity.

Intracellular Protein Transport

Randy W. Schekman, Ph.D.—Investigator

Dr. Schekman is also Professor of Biochemistry and Molecular Biology at the University of California, Berkeley, and Adjunct Professor of Biochemistry and Biophysics at the University of California, San Francisco. As a graduate student, he studied the enzymology of DNA replication with Arthur Kornberg at Stanford University. His current interest in cellular membranes developed during a postdoctoral period with S. J. Singer at the University of California, San Diego. At Berkeley, he developed a genetic approach to the study of eukaryotic membrane traffic. Among his awards is the American Society for Microbiology Eli Lilly Award in Microbiology and Immunology. Dr. Schekman was recently elected to the National Academy of Sciences.

RESEARCH in our laboratory is devoted to molecular description of the processes of polypeptide translocation from the cytosol into the endoplasmic reticulum (ER) and of vesicular transport among organelles of the secretory pathway.

Genetic and Biochemical Dissection of the Secretory Process

A genetic approach to the study of eukaryotic protein transport involved the isolation of conditional mutants. We isolated a series of secretory (*sec*) mutants in the yeast *Saccharomyces cerevisiae* that are temperature-sensitive for cell surface growth, division, and secretion. Most of the mutants accumulate secretory proteins in an intracellular pool that can be released when cells are returned to a permissive temperature. More than 30 gene products have been implicated in the process of delivering membrane and secretory proteins to the cell surface.

A combined genetic and cytologic evaluation of the *sec* mutants has allowed a description of the secretory pathway. Protein transport in yeast appears to be mediated by the same organelles and proteins that operate in mammalian cells. Molecular cloning analysis of *SEC* genes has revealed striking structural and functional homology with corresponding mammalian genes.

We have developed biochemical assays that measure the early events of polypeptide translocation into the ER and of vesicle-mediated protein transport from the ER to the Golgi apparatus. The cell-free reactions represent physiologically meaningful processes. Extracts prepared from mutant cells reproduce the temperature-sensitive defects observed *in vivo*. In favorable circumstances the mutant defects are repaired by addition of a protein fraction obtained from wild-type yeast, and such restoration of transport activity may be used to purify functional forms of *SEC* gene products.

Protein Translocation Across Membranes

Protein translocation into the lumen of the endoplasmic reticulum represents the initial step in assembly of the eukaryotic cell surface. This process has been reconstituted with detergent-solubilized membrane proteins and purified cytosolic proteins, yet the mechanism of polypeptide penetration is unclear. We have isolated mutants that are defective in translocation, using a genetic selection that requires secretory polypeptides to be retained in the cytosol. The work on these mutants is supported by a grant from the National Institutes of Health and will not be described here.

An independent line of investigation concerns the mechanism of protein translocation from the cytosol into the vacuole. The vacuole contains an array of hydrolytic enzymes and is believed to play an important role in the degradation of cytosolic proteins and intracellular membranes. The mechanism for importing substrates into the vacuole has not been evaluated.

Last year we reported a novel pathway for the localization and degradation of fructose 1,6-bisphosphatase (FBPase), a key regulatory enzyme of gluconeogenesis. FBPase is localized to the cytosol when cells are grown on a poor carbon source, such as ethanol. When cells are transferred to glucose, FBPase is degraded in a process called catabolite inactivation, which depends on active vacuolar proteases. In a protease-deficient strain, FBPase enters the vacuole and remains intact. Import into the vacuole was shown to depend on protein synthesis during the period of transfer to glucose medium. In addition, vacuolar import requires the transfer of a protein, possibly an import receptor, via the secretory pathway.

The mechanism of this new import pathway is being pursued by the isolation of mutants defective in the degradation of FBPase. Thus far a large number of genes have been identified that are required for import of FBPase into the vacuole. Preliminary evidence suggests that the import

may be mediated by an independent organelle, the peroxisome. Mutations that block peroxisomal assembly also block FBPase degradation. Furthermore, a peroxisomal enzyme, acyl-CoA thiolase, is subject to the same glucose-stimulated, vacuolar-dependent degradation process. One possible explanation of these results is that FBPase may be imported into or somehow associated with peroxisomes and then localized to the vacuole by autophagic uptake of the peroxisome.

Vesicle Transport Early in the Secretory Pathway

Subsequent stages in the secretory pathway involve protein sorting and transport from the endoplasmic reticulum to the Golgi apparatus and from there to the cell surface. Genes required for each of these steps are being evaluated by molecular cloning and by development of cell-free reactions that measure individual steps in the transport process. An assay that depends on Sec proteins has been reconstituted *in vitro*. Yeast α-factor precursor is translocated into the ER lumen of gently lysed yeast spheroplasts. In the presence of soluble proteins and ATP, the precursor is transferred to the Golgi apparatus. This system allows the purification and functional characterization of Sec proteins.

Transfer of secretory proteins from the ER to the Golgi apparatus is mediated by small vesicle carriers. In this limb of the pathway, defective *sec* mutants fall into two categories: class I, mutant cells that accumulate ER tubules at a restrictive temperature (*SEC12*, *-13*, *-16*, and *-23*); and class II, mutant cells that also accumulate several thousand 60-nm vesicles (*SEC17*, *-18*, and *-22*).

Genetic epistasis tests indicate that class I genes must execute their function prior to class II genes. This implies that class I products participate in the production of the 60-nm vesicles that are consumed, by fusion with the Golgi apparatus, through the action of class II gene products. Genetic interactions among members of class I and II genes suggest that the Sec proteins in each group act in a complex, or at least in a concerted manner, to perform their respective roles in vesicle budding or fusion.

Transport of the α-factor precursor *in vitro* is mediated by diffusible vesicles. Transport vesicles contain a core-glycosylated precursor and are physically separable from donor ER and target Golgi membranes. Budding of vesicles from the ER requires a crude cytosol fraction, ATP, Sec12p, Sec13p, Sec23p, GTP, and a *ras*-like GTP-binding protein, Sar1p. Fusion of the vesicles to the Golgi compartment is measured by conversion of the precursor to a more highly glycosylated form. Enriched transport vesicles target to the Golgi compartment and then fuse in distinct subreactions that require cytosol, Ca^{2+}, ATP, and only a subset of Sec proteins.

To allow the purification of functional Sec23p, we have developed an assay based on restoration of transport in *sec23* lysates by wild-type protein fractions. The purified protein has been isolated in two forms: a Sec23p monomer, and a heterooligomer that consists of Sec23p and a 105-kDa protein (Sec24p) that is also required for vesicle budding. The *SEC23* and *-24* genes have both been cloned, and though the sequences confirm the observed molecular weights of the isolated polypeptides, no homology to known proteins was found.

Localization of Sec proteins in yeast has met with only limited success. Yeast morphologic analysis is limited by the small cell size and the difficulty of specimen preservation. Fortunately the functional and structural conservation of Sec proteins allows their localization in mammalian cells. The first and most favorable example is Sec23p, where antibodies directed against the yeast protein cross-react with the mammalian homologue. Immunolocalization studies reveal a striking enrichment of Sec23p in the cytoplasmic pocket that separates the transitional ER and the cis face of the Golgi complex. This location is completely consistent with the role proposed for yeast Sec23p and further confirms that the secretory pathway is fundamentally conserved across the broad spectrum of the eukaryotic kingdom.

Development and Function of the Synapse

Richard H. Scheller, Ph.D.—Associate Investigator

Dr. Scheller is also Associate Professor of Molecular and Cellular Physiology and Associate Professor of Biological Sciences (by courtesy) at Stanford University. He received his B.S. degree from the University of Wisconsin–Madison and his Ph.D. degree in chemistry from the California Institute of Technology. Dr. Scheller was a postdoctoral student with Richard Axel and Eric Kandel at Columbia University.

THE nervous system is composed of large numbers of unique cells that communicate with each other via the regulated release of chemical neurotransmitters. These synaptic interactions govern animal behavior, and modulation of the efficacy of synaptic communication is thought to underlie learning and memory. We are interested in understanding the molecular mechanisms of synaptic formation during development and regeneration in the peripheral nervous system after nerve injury. It is also our goal to contribute to an understanding of how the nerve terminal functions in the regulation of neurotransmitter release.

Processing and Packaging of Neuropeptides

Many synapses release two types of chemical messengers: fast-acting, or classical transmitters, and slower-acting messengers, or neuromodulators. Most of the various chemicals used as messengers in the brain are neuropeptides. These molecules are synthesized as larger precursors that are processed to smaller active peptides. One interesting neuropeptide precursor is expressed in an identified set of neurons, the bag cells, in the marine snail *Aplysia*. When these neurons fire, they release a set of neuropeptides derived from a single precursor. These peptides act on neurons and peripheral tissues to regulate egg laying, a stereotyped behavior.

Interestingly, the peptides produced on the egg-laying hormone (ELH) precursor are packaged in two types of vesicles. These vesicles contain different sets of peptides and are differentially localized within the neurons. We are interested in understanding how the peptides initially synthesized on a single precursor are sorted into different vesicles. We are also interested in understanding the physiological significance of the differential packaging and localization.

When the ELH precursor is transfected into mammalian pituitary tumor cells (AtT-20 cells), ELH is packaged with the endogenous hormone. The amino-terminal region of the precursor is degraded within the secretory cell, probably in the lysosomes. Thus the AtT-20 cells, like the bag cells, differentially route the two regions of the ELH prohormone. Mutating the first cleavage site from a set of four basic residues to two basic residues results in constitutive secretion of the amino-terminal region of the precursor, not intracellular degradation. These results suggest that the subcellular location of the first endoproteolytic processing event is critical in determining the routing of the processing intermediates.

Mechanisms of Synaptic Transmission

When the action potential travels down the nerve and enters a release zone, changes in the membrane potential open channels that allow calcium to enter the cell. The calcium promotes transmitter release and membrane fusion. The membrane then recycles, forming new vesicles, which are then replenished with chemical transmitter. This cycle might be considered the fundamental process that underlies nervous system function, yet little is known about the molecular mechanisms involved. In an attempt to define the molecular mechanisms that regulate membrane flow in the nerve, our laboratory and others have begun to characterize the proteins associated with the critical organelle in the process, the synaptic vesicle. For these studies, we use mammalian brain and the electric organs of marine rays. These electric organs have a concentration of synapses approximately 100-fold higher than that of skeletal muscle. In addition, these synapses are homogeneous; they all use the neurotransmitter acetylcholine.

Purified synaptic vesicles contain about 20–50 protein bands when fractionated on acrylamide gels. Genes encoding many of these proteins have been characterized and the primary sequence of the molecules determined. Some of the proteins show interesting homologies to other molecules, and others are turning out to have counterparts in yeast where genetic studies of membrane trafficking have provided insight into the secretory process. It has also become apparent that many of the synaptic vesicle proteins are members of small gene families. Individual members of these gene families are differentially expressed through the

brain, resulting in a variety of combinations of these molecules on different vesicles.

Molecular fractionation and immunoprecipitation techniques suggest that several of the synaptic vesicle proteins interact to form a large multimeric complex. This complex contains a previously uncharacterized 35-kDa protein that is a major substrate for casein kinase. We have purified this protein and isolated cDNAs encoding the molecule. The protein is unique in the database and is predicted to be anchored in the membrane by a hydrophobic carboxyl terminus and to have an amino-terminal region oriented toward the cytoplasm. Like the synaptic vesicle proteins, this molecule is a member of a small gene family. Experiments are directed toward understanding the localization and function of this protein and the other synaptic vesicle proteins.

Synapse Development

Motor neurons in the spinal cord send axons to muscle fibers throughout the body. When axons contact muscle fibers, a highly ordered structure consisting of a presynaptic nerve terminal and a postsynaptic site develops. The presynaptic terminal comprises an active zone rich in synaptic vesicles containing neurotransmitter. The postsynaptic element is made up of a membrane rich in receptors for the neurotransmitter and an indentation in the membrane called the junctional fold. An extracellular matrix, or basal lamina, surrounds the muscle fiber, including the space between the nerve and muscle.

One of the key events in the development of the neuromuscular junction is the redistribution of neurotransmitter receptors that occurs when nerve contacts muscle. Initially receptors for the neurotransmitter, in this case acetylcholine, are randomly distributed on the muscle fiber. When the nerve contacts muscle, neurotransmitter receptors aggregate under the nerve terminal in an appropriate position to detect the chemicals released during synaptic transmission.

Agrin, a component of the extracellular matrix, causes acetylcholine receptors to cluster when added to muscle fibers growing in culture. We have isolated recombinant DNA clones encoding agrin molecules and, through an analysis of the nucleotide sequence, have defined the primary amino acid sequence of the molecule. When we compare the predicted agrin sequence with the proteins in the data bank, two types of similarities are revealed. The first is to a class of molecules that inhibit proteases and the second to a protein motif called EGF (epidermal growth factor) repeats. The gene is expressed in embryonic motor neurons at the time they are first contacting muscle fibers. Two regions of the agrin gene are alternately spliced and may produce up to eight forms of the molecule.

Expression of agrin encoding cDNAs in CHO and COS cell lines results in the association of the protein with the surface of the transfected cells, probably through assembly into an extracellular matrix. Coculture of agrin-expressing cells with primary muscle fibers or a C2 myoblast cell line results in aggregation of acetylcholine receptors at sites of contact between the transfected cell and the muscle fibers. Only the forms of the protein containing an eight–amino acid sequence, which is the product of alternate RNA splicing, are capable of causing clusters on S27 cells, a mutant C2 line lacking proteoglycans. These data suggest that agrin may cluster receptors via two mechanisms, one of which is proteoglycan dependent. Another possibility is that the eight amino acids provide a binding site that results in a high-affinity interaction, overriding the need for the proteoglycan component.

Since agrin is stably maintained in the synaptic basal lamina after nerve or muscle damage, it may also play a role in regeneration events. Understanding the mechanisms of peripheral synapse regeneration may lead to procedures that could aid in central nervous system regeneration.

Molecular Pathogenicity Studies of Enteric Bacteria

Gary K. Schoolnik, M.D.—Associate Investigator

Dr. Schoolnik is also Associate Professor of Medicine and of Microbiology and Immunology at Stanford University School of Medicine. He received his M.D. degree from the University of Washington. He was an intern, resident, and chief resident in internal medicine at Massachusetts General Hospital, a fellow in infectious diseases with King Holmes and Thomas Buchanan at the University of Washington, and a postdoctoral fellow with Emil Gotschlich at the Rockefeller University. He founded the Division of Geographic Medicine at Stanford University and has established a research center for the study of infectious diseases in southern Mexico.

BACTERIAL, viral, and parasitic infections of the gastrointestinal tract cause an estimated 500 million illnesses and 5 million deaths each year among children living in the developing countries. The principal mission of our laboratory is to discover how these infectious agents cause disease, how they are spread, and how this information can lead to new tactics for prevention and control. This effort has entailed work in two settings: molecular studies in our laboratory at Stanford University, and epidemiological investigations at a field laboratory in southeastern Mexico, where infections of this kind are common.

In the first setting, the unit of analysis is the organism itself. We try to determine how it attaches to, invades, and damages human cells. This anthropomorphic orientation views the disease-causing capacity of an infectious agent as the central research issue. In the second setting, the unit of analysis is a household or a village. In this context we seek to understand how the organism is transmitted within the community, what its reservoirs are, and how it manages to survive as a viable entity in the real world.

This ecological orientation seeks to understand how the organism adapts to different environmental habitats. Within the context of a Third World village, these habitats include contaminated food and well water, sewage, and the gastrointestinal tracts of people and animals. By using biochemical and genetic tools to study infectious agents as they inhabit and move among these different environmental niches, we have begun to understand the underlying molecular mechanisms for this remarkable capacity. This in turn is beginning to lead to new strategies for the control of these diseases through vaccination and epidemiological interventions. Examples of work in progress are described below.

Enteropathogenic *Escherichia coli*

Enteropathogenic *E. coli* (EPEC) are a common cause of infantile diarrhea in Third World children. When biopsies of the small intestine are performed, colonies of EPEC are found attached to the underlying epithelia. It is evident that the bacteria interact not only with the host cells to which they are bound but also with each other. Beneath these adherent colonies, structural changes in the host cell also occur, indicating that EPEC have altered the absorptive surface of the intestinal cell. It is now clear that structural changes of this kind are directly responsible for the diarrheal syndrome.

From these studies we have learned that EPEC's pathogenic strategy consists of at least three distinct steps, which probably occur in the following order: the coalescence of individual bacteria into infectious units composed of several to hundreds of organisms (an event that probably occurs in the small intestine, soon after ingestion of the bacteria); attachment of these infectious units to the surface of intestinal epithelial cells; and, following close contact between the bacteria and the epithelial cells, the organism's partial penetration of the cell, in association with concomitant rearrangement of its cytoskeleton.

Each of these steps is performed by a distinct surface structure of the bacteria. First, the bacteria are organized into infectious units through the production of rope-like appendages that emanate from the organism's surface. Termed "bundle-forming pili," these appendages create a network within which the bacteria become enmeshed. Then the bacteria bind to human intestinal epithelial cells through the activity of rod-like filaments that project from the organism like spines from a porcupine. These filaments bind fibronectin molecules that are located around the periphery of each intestinal epithelial cell. Finally, an outer membrane protein of the organism interacts with integrin-like molecules of the epithelial cell, an event that triggers changes in the architecture of the cell's cytoskeleton.

The production of at least one of these surface structures is controlled by physiochemical signals that operate in the intestinal lumen, where they serve as unique signatures of that habitat. Thus, when under the influence of these signals,

EPEC produce bundle-forming pili and coalesce as infectious units. In contrast, when in inanimate environmental reservoirs such as water or sewage, their production is repressed and they exist as single, well-separated organisms. This phenomenon, together with the findings described above, shows that EPEC exemplify two features of bacterial pathogenesis: first, that pathogenicity is often the consequence of several distinct virulence determinants of the organism acting together and, second, that the expression of these determinants may be regulated by local features of the organism's environment.

We are now attempting to use these findings to design an EPEC vaccine for the prevention of diarrheal illnesses in children. We hope to delete from the EPEC chromosome the gene that causes functional abnormalities of the host cell while retaining the genes that direct the cell attachment capacity and the production of bundle-forming pili. A genetically modified strain of this kind should be attenuated with respect to its virulence, yet fully capable of colonizing the small intestine, where it would be expected to stimulate a protective immune response.

Using Attenuated Strains of the Typhoid Fever Bacillus as Vaccines

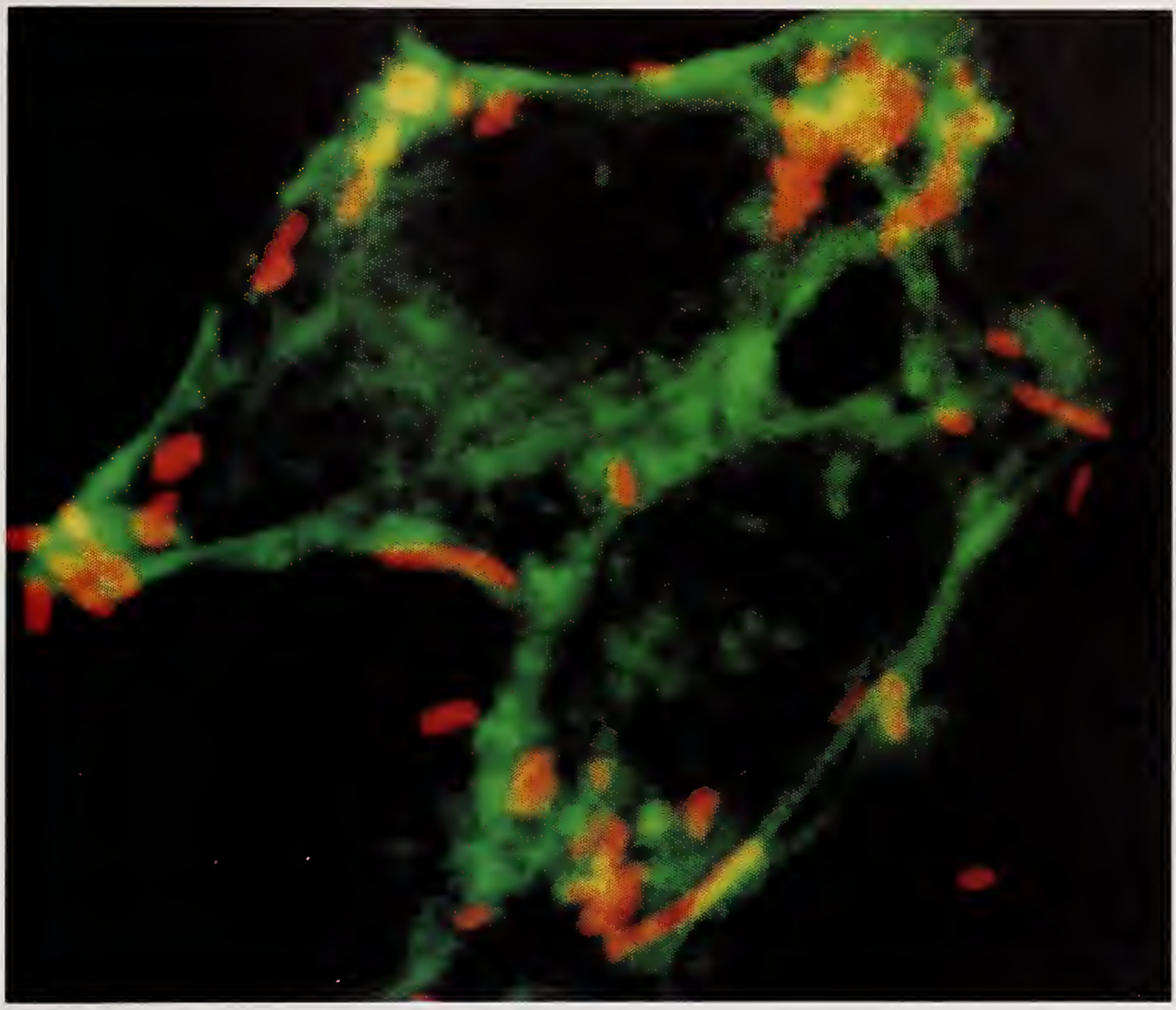
Fully virulent strains of *Salmonella typhi* cause typhoid fever, whereas an attenuated strain of the same species, administered as a living oral vaccine, is now FDA approved and widely used for the prevention of typhoid fever. Attenuated strains of this kind have been proposed as vehicles that might be able to carry other vaccine substances—termed "passenger antigens"—derived from a variety of microorganisms. Such a vaccine would stimulate immunity not only to *S. typhi*, but to other infectious agents as well. We have been studying this system for the delivery of T cell epitopes.

For the purposes of this study, T cell epitopes are considered to be localized regions of a protein that stimulate a cellular immune response and are required for the elimination of intracellular pathogens, including some bacteria and many viruses and protozoan parasites. We have found that T cell epitopes can be effectively processed and presented to the immune system when expressed in the center of the flagellin protein of *S. typhi*. Flagellin is the protein building block of a whip-like extracellular filament that acts like a motor to propel the bacterium through liquids. Because these vaccine strains would express a flagellin protein that would also contain a foreign T cell epitope, they are referred to as "chimeric" flagellins.

We have taken advantage of recent information about how epitopes of this kind are processed by cellular components of the immune system. It is now clear that one common pathway would entail the presence of the chimeric flagellin in the endosome of a macrophage. In that specialized compartment of the cell, it would be digested by endosomal proteolytic enzymes, resulting in the release of the epitope as a small peptide prior to its presentation to other components of the immune system. A T cell epitope was placed in the flagellin protein flanked at either end by amino acids known to be sites of cleavage by endosomal proteolytic enzymes. This has yielded a five- to eight-fold increase in the magnitude of the resulting immune response.

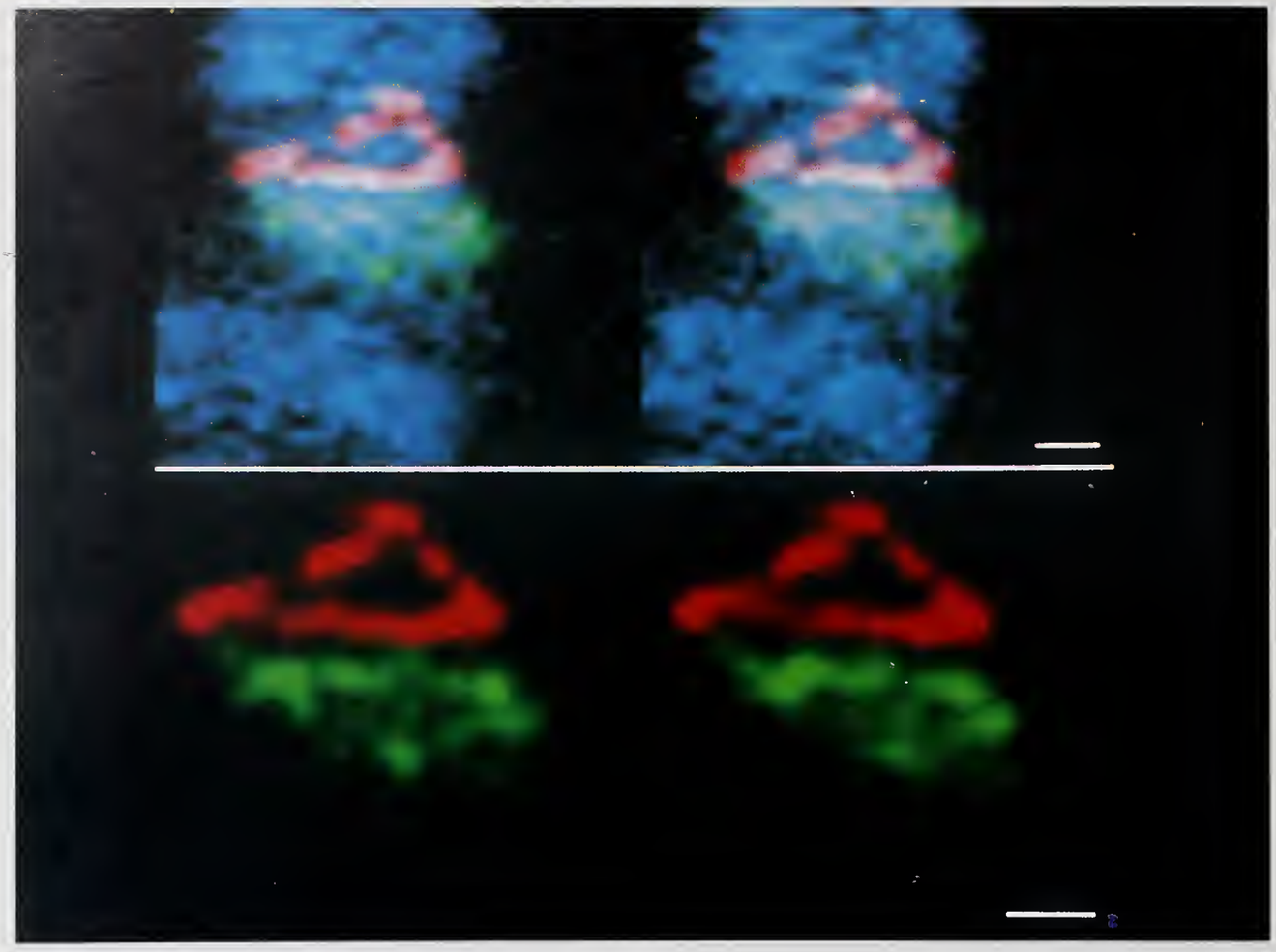
Further modifications have entailed placing the expression of these chimeric flagellins under the control of the heat-shock protein promoter, which is known to be activated in *Salmonella* growing in macrophages. Other modifications include the use of multiple tandem copies of these epitopes and the use of flagellins that are secreted as soluble proteins into the macrophage endosome. Taken together, these studies have demonstrated that it is possible to use attenuated strains of *S. typhi* to deliver T cell epitopes and that the effectiveness of this delivery system can be enhanced by the design features described above.

We are now working in collaboration with scientists at the National Institutes of Health to place T cell epitopes of a human immunodeficiency virus (HIV) protein in *Salmonella* flagellins, with a view to preparing useful vaccines for the prevention of AIDS.



HEp-2 cells infected with Escherichia coli *express the* inv *gene from the invasive bacterial pathogen* Yersinia enterocolitica. *Data were recorded from a computer-controlled confocal laser microscope and processed on a graphics display terminal. The HEp-2 cells (green) were stained with the actin-specific stain FITC-phallicidin, while the invasin-bearing bacteria (red) were visualized by indirect immunofluorescence using antisera against* E. coli *outer-membrane proteins. Areas of extensive overlap appear yellow. Results indicate that actin, a major component of the cytoskeleton that controls cell shape and movement, accumulates in the region of entering bacteria.*

From Young, V.B., Falkow, S., and Schoolnik, G.K. 1992. J Cell Biol *116:197–207, by copyright permission of the Rockefeller University Press.*



Stereo pairs showing a three-dimensional view of in situ *hybridization signals to two neighboring genes. DNA from the* white *gene (seen here in green) and from the* Notch *gene (red) were hybridized to the giant chromosomes from* Drosophila *salivary glands, fluorescently labeled, and visualized by wide-field three-dimensional optical sectioning microscopy. Top picture shows these signals overlaid on the chromosome (blue); bottom, the signals alone. The structure of the two genes is seen to be distinctly different. Scale bar = 1 μm.*

Research and photograph by Susan Parmelee in the laboratories of John Sedat and David Agard.

Three-Dimensional Structure of Eukaryotic Chromosomes

John W. Sedat, Ph.D.—Investigator

Dr. Sedat is also Professor of Biochemistry and Biophysics at the University of California, San Francisco. He received his Ph.D. degree in biology from the California Institute of Technology. His postdoctoral work with Fred Sanger was done at the Medical Research Council in Cambridge, England. Before joining the faculty at UCSF, Dr. Sedat was a research associate at Yale University.

THE three-dimensional structure of chromosomes, both in the nucleus and during cell division, remains a major unsolved problem in biology. Our laboratory, in collaboration with that of David Agard (HHMI, University of California, San Francisco), is investigating chromosome structure from the perspective of several interlocking questions: 1) What is the architecture of the chromosome in the intact diploid nucleus? How does the three-dimensional structure change as a function of development, or progression through the cell cycle? 2) What is the architecture of a given gene in the nucleus? Do the structural attributes reflect the detailed molecular information? 3) How do interphase chromosomes condense to form the intricate mitotic structure at cell division?

The fruit fly *Drosophila melanogaster*, well-known for its genetics, development, and biochemistry, was chosen as a model biological system. Although the initial emphasis is structural, molecular genetics and biochemistry provide functional correlations.

The UCSF three-dimensional optical microscope has been developed to the point that data at several wavelengths can be routinely collected, even as a function of time (four-dimensional microscopy), and can be used without computer experience. Still, we continue to perfect and enhance the instrumentation. We increased the time resolution for data collection and greatly improved the image quality of the four-dimensional data, permitting analysis of much information on biological structures. We continue to write software, with extensive mathematical analysis, to correct systematic image acquisition problems, to display results in a variety of formats, and to model and analyze, often quantitatively, the complex three-dimensional data. We have started to develop a computer-based methodology to extract and analyze quantitatively the large- and small-scale motion of chromosomes or structure within the nucleus.

Four-Dimensional Optical Microscopy

We have continued to study the structure of the cellular nucleus in living *Drosophila* embryos. Nuclei were labeled by microinjection of fluorescent histones, or other chromosomal proteins. Nuclear and chromosomal structures were followed throughout the cell cycle during embryonic development. In addition to discerning structural changes, we can now infer function.

Topoisomerase II—A Key Nuclear Protein

Our studies include an effort to understand the role of various proteins in the organization and dynamics of chromosomes. We have therefore studied the distribution and dynamics of the DNA strand-passing (unknotting) enzyme topoisomerase II. High-resolution three-dimensional imaging of *Drosophila* embryonic chromosomes shows a heterogeneous distribution of topoisomerase II along the chromosome. At metaphase and anaphase, the enzyme can be clearly seen to be situated adjacent to the chromosome. This suggests that its localization may be linked to its activity: the enzyme may concentrate at sites of chromosome condensation and/or segregation. These data argue against a purely structural role for the enzyme.

We have studied the dynamics of localization by injecting fluorescently labeled antibodies against the enzyme, or the labeled enzyme itself, into live embryos and then imaging them by our three-dimensional microscopy as a function of time. The resulting time-lapse movies have shown that the concentration of nuclear topoisomerase II changes dramatically throughout the cell cycle. The highest levels occur in late interphase, the lowest levels in telophase. A very dynamic fibrillar complex is evident during interphase. Experiments that will disclose the functional relevance of this structure are in progress.

A Molecular Dissection of the Nuclear Periphery

Recently we showed that the lamin proteins of the nuclear envelope (NE) form a highly discontinuous network in somatic interphase nuclei. Several obvious questions arise. First, where are

the other known components of the nuclear periphery (pore complexes, chromatin) relative to this network? Second, what, if anything, occupies these large, lamin-empty regions? Third, how are these structures assembled as the NE re-forms during telophase?

Chromatin in the nuclear periphery displays an interesting structural paradox in that a large fraction appears to be aligned beneath the lamin network, but with very little contacting lamins directly. The majority of it seems to be at a distance of about 0.2 μm. This result is consistent with much indirect evidence for a strong interaction between chromatin and the nuclear lamina, but strongly suggests that a direct physical contact is not involved.

We injected lamins and lamin-specific monoclonal antibody Fab fragments, both fluorescently labeled, into early *Drosophila* embryos to study four-dimensional lamin-NE dynamics. In these experiments the embryos develop normally and hatch on time. We observe a highly discontinuous lamin network *in vivo*, with interlamin fiber spacings at least as large as those observed in fixed samples. From these experiments, new four-dimensional data, spanning prophase to metaphase in the cell cycle, show a surprisingly complex series of lamin structural rearrangements. Lamins do not completely disassemble and disperse at the onset of mitosis, but remain well localized with complex structural dynamics until well into the mitotic process. Further lamin structural changes take place at anaphase and telophase.

These studies emphasize that lamins appear essential for the nuclear structural reorganizations that take place at all points of the cell cycle. If, however, we inject fluorescently labeled interphase lamins, a very different picture results. Arrested nuclear structures leading to chromosomal/nuclear aggregates are seen. These studies suggest that structural/functional assays will be required for proper interpretation of the biochemistry.

Three-Dimensional *in situ* Hybridization

We are continuing our study of nuclear organization, using three-dimensional fluorescence *in situ* hybridization. The studies, described in our previous report, were performed in *Drosophila* embryos, primarily at the histone gene locus, a 500-kb, tandemly repeated gene cluster. We have extended our techniques to other whole-mount tissues from *Drosophila*, including developing imaginal tissues (which will later form the adult fly). We have also improved our hybridization protocols and can now detect chromosomal probes as small as 12 kb with high signal-to-noise resolution. Using these techniques, we are investigating the arrangement of chromosomes throughout the cell cycle and as a function of development.

A focus of our studies is the question of homologous chromosome pairing (one chromosome from the male parent, the other from the female). This is of particular interest in *Drosophila* biology because genetic evidence has indicated that different alleles at certain homologous loci can influence one another, implying that the loci communicate in some fashion. Such phenomena have been termed transvections, or more generally, "trans-sensing effects." By determining the nuclear positions of the histone gene cluster in developing embryos, we have shown that the homologous chromosomes bearing this locus are for the most part spatially distinct, or unpaired, throughout most of early embryonic development, but that a transition occurs just prior to cellularization, a distinct time point in development, after which the locus is seen to be paired at high frequency.

We are currently extending this analysis to other genetic loci, to determine whether this unpaired/paired transition occurs simultaneously at all chromosomal positions. We have preliminary evidence that loci more distant from the centromere than the histone locus may show different pairing behavior in early embryos. We are particularly interested in carrying out this type of analysis for loci that are known to have transvection or trans-sensing effects.

This general methodology has a number of potential applications to problems of cell lineage, neural architecture, and pattern formation in development. We are pursuing some of these interests in collaboration with other laboratories.

A Molecular Basis of Familial Hypertrophic Cardiomyopathy

Jonathan G. Seidman, Ph.D.—Investigator

Dr. Seidman is also Professor of Genetics at Harvard Medical School. He received his undergraduate degree from Harvard University and his Ph.D. degree from the University of Wisconsin–Madison, where he studied with William McClain. His postdoctoral studies were carried out in Philip Leder's laboratory at the National Institute of Child Health and Human Development.

FAMILIAL hypertrophic cardiomyopathy (FHC) is a heart muscle disorder with an autosomal dominant pattern of inheritance. The disease is characterized clinically by unexplained myocardial hypertrophy and variable symptomatology that can include syncope, arrhythmias, congestive heart failure, and sudden death. Diagnosis in young people is particularly important. The incidence of sudden death appears higher in this group and can occur without warning.

Indeed, hypertrophic cardiomyopathy is one of the most common autopsy findings among young athletes who die suddenly. The large majority of these were undiagnosed previously. Diagnosis in this age group may be particularly difficult, since the diagnostic clinical and echocardiographic criteria may not be manifest until adulthood.

Because mutations in the cardiac myosin heavy-chain (MHC) genes were implicated as the cause of FHC in two families, we decided to undertake a direct analysis of these genes in affected individuals from other families. During the past year we spent considerable effort attempting to find in 24 unrelated families the cardiac MHC mutations that cause FHC.

We had previously demonstrated that most FHC mutations are missense or point mutations in the MHC genes. We have successfully employed a variety of techniques in the detection of missense mutations within genes. Most of these are based upon amplification of genomic DNA sequences and analyses of individual exons. Application of these approaches to the study of FHC mutations was more difficult because the β MHC polypeptide is encoded in 40 exons, and hence 40 independent analyses are required to examine the entire gene. Furthermore, FHC is an autosomal dominant disorder, and affected individuals are heterozygous, bearing one mutated and one normal gene. Genomic analyses may fail to detect deletions of entire exons or mutations that altered gene splicing because of the presence of one normal gene.

Access to messenger RNA in which intronic sequences have been excised would overcome these limitations and allow analysis of coding regions in a more rapid and convenient manner. Although MHC mRNAs are abundant in the heart, expression elsewhere is low and restricted to selected fibers in slow-twitch skeletal muscle. Normal and mutant β MHC sequences were detected in RNA transcripts from peripheral lymphocytes and lymphocyte cell lines transformed by Epstein-Barr virus (EBV). This finding permitted examination of β MHC mRNA, even though cardiac tissue was not available.

Seven different β cardiac MHC mutations were found among 24 unrelated FHC probands. Four mutations were identified in two or more families. One of these, the Arg453Cys mutation, probably occurred independently in families B and E, because only family B also contains a hybrid α/β cardiac MHC gene on the same chromosome. Whether other mutations that are shared by apparently unrelated individuals arose independently within mutational hot spots or represent a founder mutation is uncertain. Characterization of other families should also elucidate whether the mutations that can cause the FHC phenotype are restricted in number.

The identification of seven different mutations in a disease with significant morbidity and premature death suggests that many of these are of relatively recent origin in human evolution. Since FHC mutations are not likely to provide a selective advantage, but have not been lost in the population, they probably reflect a high incidence of new mutational events in the β cardiac MHC gene.

The natural history of FHC is quite variable, and diagnostic tests have been unable to identify those with a more serious prognosis or those at risk for sudden, unexpected death. To determine whether particular β cardiac MHC gene mutations correlate with clinical outcome, we compared several indices with genotype. Data from families with the same mutation are pooled. Disease-related deaths were infrequent in families with the Val606Met mutations as compared with the Arg249Gln, Arg403Gln, or Arg453Cys mutations. While the incidence of disease-related deaths in individuals with the Arg249Gln muta-

tion is similar to that associated with other mutations, the average age at death is significantly older for affected individuals within this family.

To assess cumulative survival of affected individuals with respect to age, we compared survival curves of FHC families with five different mutations for which sufficient numbers of affected individuals (alive or deceased) were available. These analyses confirmed that persons with the Val606Met mutation survive longer than those with the Arg453Cys or Arg403Glu mutations.

The Arg249Glu mutation appears to produce an intermediate phenotype. Survival in these individuals is better than in those with the Arg453Cys or the Arg403Glu mutation. While survival appears shorter in individuals with the Arg249Glu mutation than in those with the Val606Met mutation, this difference is not statistically significant. Individuals with the Arg453Cys mutation (with or without the hybrid gene) and those with the Arg403Cys mutation have similar life expectancies, dying prematurely.

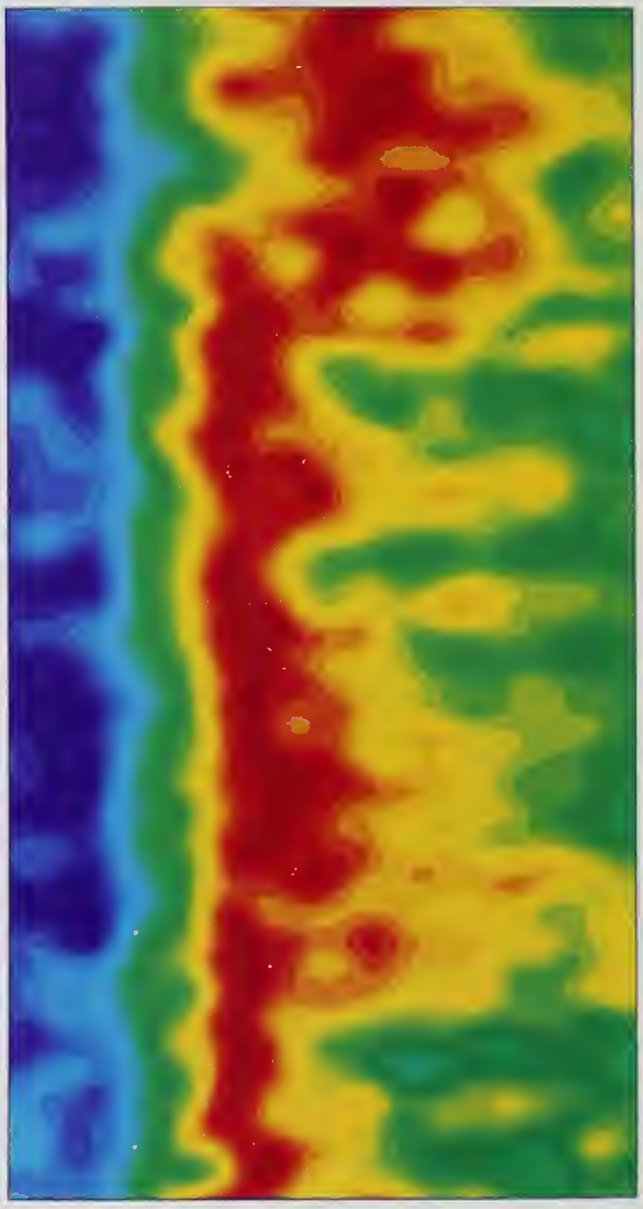
The seven different mutations are clustered in the globular head of the polypeptide, and we postulate that the defective myosins made by these genes poison myosin function by impairing physiologic interactions with other contractile elements. Six of seven missense mutations affect the charge of the altered residue. Perhaps the fact that the Val606Met mutation does not alter the net charge of the polypeptide accounts for the better survival of affected individuals.

Identification of multiple disease-causing mutations implies that mutational events within the β cardiac MHC gene are not uncommon and that a number of FHC-causing cardiac MHC mutations have occurred during the course of human evolution. A high frequency of myosin mutation may explain the relatively high incidence of sporadic (10–20 percent) hypertrophic cardiomyopathy.

The value of identifying FHC mutations was further demonstrated by analysis of a large family affected by mutation Arg249Glu. Related adult family members were clinically evaluated for FHC, and blood samples were obtained for independent genetic diagnosis. The clinical and genetic diagnoses were in complete concordance.

Thirteen children (ages 2–20) were also evaluated. Statistically, half should have been affected, but only one child had clinically demonstrable FHC. Genotype analysis of these children revealed six who inherited the mutant MHC gene.

These data underscored the insensitivity of clinical diagnostic criteria for FHC in children and young adults. Genetically based diagnoses of FHC permit preclinical diagnosis and should facilitate prenatal diagnosis. Furthermore, the ability to make a preclinical diagnosis in families makes possible longitudinal studies of disease development and interventional trials.



Time course of the calcium concentration in the dendrite of a CA1 hippocampal pyramidal neuron in response to synaptic stimulation as a function of time, running horizontally. The concentration along the dendrite, running vertically, is shown in false color, with the resting level in blue and the highest concentration in red.

Research of Richard Adams and Terrence Sejnowski.

Computational Neurobiology of Sensory Representations

Terrence J. Sejnowski, Ph.D.—Investigator

Dr. Sejnowski is also Professor at the Salk Institute for Biological Studies and Professor of Biology and Neuroscience at the University of California, San Diego. He received his B.S. degree in physics from Case Western University and his M.A. and Ph.D. degrees in physics from Princeton University. He was a postdoctoral fellow with Alan Gelperin in the Biology Department at Princeton and with Stephen Kuffler at Harvard Medical School, where he studied mechanisms of synaptic transmission. Dr. Sejnowski was a member of the faculty of the Biophysics Department at the Johns Hopkins University before moving to San Diego. He and Patricia Churchland have recently written The Computational Brain, *a book on computational neuroscience.*

WE do not yet understand how the nervous system enables us to recognize objects, to learn new skills, and to plan actions. The discovery that single neurons in the visual system can be highly selective in responding to visual stimuli led to the view that the perception of complex objects could be directly linked to the activity of individual neurons. This possibility raises a number of questions, such as what degree of influence a single neuron can have on behavior and whether there are enough neurons in the brain to account for the large number of objects that can be perceived.

An alternative possibility relies on populations of neurons to represent perceptual states. On this account, the information essential to the representation of an object is distributed over a large population of neurons. It is difficult to imagine how a pattern of activity in a large number of neurons distributed widely throughout the brain could be used to recognize an object and serve as the input for motor actions. Computer models incorporating cellular information from single-cell recordings and constrained by psychological measurements on performance can help to organize these data and provide a conceptual framework for understanding distributed representations. Such models are being used to explore how the visual cortex represents the three-dimensional world, how this representation may arise during development, and how the information coded by these neurons might be used to coordinate actions such as eye movements.

The perception of depth depends upon a number of visual cues, but only one of them relies on the slight positional shift that occurs between the different viewpoints of the two eyes, called the image disparity. Neurons in the first cortical stage of vision are sensitive to disparity, and the development of this sensitivity is dependent on binocular vision during a critical period.

In the adult visual cortex, neurons are observed to be either dominated by input from one eye (monocular cells) or relatively balanced with input from both eyes (binocular cells). Furthermore, binocular cells tend to be stimulated maximally when images are in exact correspondence in both eyes, thus preferring zero disparity, and relatively monocular cells tend to prefer nonzero disparities. We have simulated the development of a layer of cortical cells receiving inputs from both eyes and show how such a relationship between ocularity and disparity might arise.

The key feature of our model is the use of correlations of activity both within each eye and between the eyes. We assume that two retinal cells close to each other will have more correlated activity than two cells far apart. Corresponding points in the two eyes will also tend to be correlated, since they will look, on average, at the same point in space. However, the correlation between the eyes will also be spread out by convergent and divergent eye movement. We can simulate visual development in two stages: prenatal, when the two retinae have essentially independent activities, and postnatal, when the eyes are open and have correlated activities. By varying the amount of development that occurs in the model before eye opening, we can show that a mixture of monocular and binocular cells arises with the observed relationship to disparity.

Disparity provides information about the relative positions of objects in space, but this cue is insufficient to recover the absolute distance of objects from the viewer. However, the distance of an object from the viewer can be computed by combining relative depth cues with other information, such as eye position. We have developed a network model to explore how the vergence of the two eyes (angle between the two lines of sight) and the binocular disparity could be combined to represent the distance to an object.

Single neurons have a wide range of disparity tuning curves that are broad and overlapping. Such a distributed representation of disparity was used in a network model to encode the inputs. The network was trained to transform disparity and vergence input information by projections

through a layer of hidden units to an output layer that represented the perceived egocentric depth of the object, as determined by psychophysical measurements. The disparity tuning curves of the hidden units were similar to those of the input units, and varying the vergence did not change the shape of the tuning curves; however, the vergence did modulate their amplitude.

Similar "gain fields" for conjugate eye movements have been observed in the posterior parietal cortex, a region of the brain that is essential for our internal representation of external space. The predictions of this model can be tested by recording single-unit activity in the cerebral cortex of awake and behaving monkeys, and several laboratories are pursuing these experiments. Preliminary results support the model.

Neurons in the early stages of visual processing in the cerebral cortex are organized in retinotopic maps. Thus visual features are arranged in a system of coordinates that is based on the position of features in the visual field of the retina rather than on the absolute position of features in space. Psychological experiments provide further evidence that simple visual features such as orientation and direction of motion are organized according to retinal coordinates. At later stages of visual processing, the receptive fields of neurons become very large; and in the posterior parietal cortex, containing areas important for sensorymotor coordination, the visual responses of neurons are modulated by both eye and head position. A previous model of the parietal cortex showed that the modulation of the neurons observed there is consistent with a distributed spatial transformation from retinal to spatial coordinates. Our model of the transformation from disparity to distance by vergence modulation can be considered a generalization of this model to include the third dimension of space.

All these models assume that the responses of neurons in the early stages of visual processing in cerebral cortex depend only on retinal information and not on the direction of gaze. Several laboratories have now reported that eye position does in fact modulate the visual response of many neurons in early stages of visual processing. Furthermore, this modulation appears to be qualitatively similar to that previously reported for neurons in the parietal cortex. These new findings suggest that transformations from retinal to spatial representations could be initiated much earlier than previously thought.

We have used network models to study the consequences of incremental spatial transformations in a feedforward hierarchy of cortical maps. Our model shows that it is possible for visual features to be encoded in spatial coordinates already at very early stages of visual processing. We call this new type of spatial map a retinospatiotopic representation and are exploring its counterintuitive properties. The model makes several surprising predictions that we are testing with perceptual experiments on human observers.

The primate visual system is very good at complex motion-processing tasks such as tracking a moving object against a textured background under a variety of luminance conditions. In order to track a moving object, the visual system must integrate many local motion estimates from many neurons, each with limited spatial receptive fields. No single neuron has the information needed to estimate the velocity of the object.

We have developed a simple model for motion processing in the visual areas of cortex that specialize in representing motion. The model assumes two pools of filters at each location on the visual field: one pool computes estimates of motion in a local region of the visual field, while the other estimates the relevance or reliability of each local motion estimate, based on the estimate itself and on additional information from the visual scene. Outputs from the second pool can "gate" the outputs from the first pool through a gain-control mechanism, before the local motion estimates are integrated to form more-global estimates. The proposed mechanism of gain control is consistent with measured responses of cortical cells under conditions of interfering motion of transparent stimuli.

These models provide representations of objects in space that are highly distributed. We also want to understand how these distributed representations can be used to direct the motor system to orient toward these objects. For example, motion estimates can be used to direct the eyes to track moving objects, and distance estimates can be used to guide hand movements to reach out for objects. We are developing models of motor systems in the brain that will complement these models of sensory processing. The models of motor control are based on networks of neurons that include feedback connections, which makes them highly dynamic. New principles of neural processing may emerge as more-detailed dynamical properties of neurons are incorporated into these models.

Adenovirus as a Model for Control of Gene Expression

Thomas E. Shenk, Ph.D.—Investigator

Dr. Shenk is also Elkins Professor of Molecular Biology at Princeton University and Adjunct Professor of Biochemistry at the Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey. He received his Ph.D. degree in microbiology from Rutgers University for studies with Victor Stollar, and his postdoctoral training with Paul Berg at Stanford University. Before coming to Princeton, he was Assistant Professor of Microbiology at the University of Connecticut Health Center and then Professor of Microbiology at the State University of New York School of Medicine at Stony Brook. Dr. Shenk counts among his honors the Eli Lilly Award in Microbiology from the American Society for Microbiology and an American Cancer Society Professorship.

ADENOVIRUSES are widespread, and humans are first infected when quite young. Generally the infection results in cold-like symptoms and resolves without complication. Some human adenoviruses, however, induce a variety of benign and malignant tumors if injected into a rat or hamster. Since these viruses contain DNA and are tumorigenic under certain conditions, they are classified as DNA tumor viruses.

Adenoviruses can be propagated easily in cultured cells. When human cells are infected, the approximately 30 viral genes are expressed, the viral chromosome is replicated, and individual DNA molecules are packaged into protein shells to produce virus progeny. Since viral genes are expressed at high levels compared with most cellular genes, and since this expression is tightly regulated, the adenovirus is a useful probe for studying the control of gene expression.

During the past year, much of our effort has focused on transcriptional control of viral gene expression. The first viral gene to be expressed after infection of a cell is the *E1A* gene, which encodes a protein that activates expression of additional viral genes. The E1A protein appears to activate transcription (copying of genetic information) through several physiologically distinct mechanisms. One of these involves a cellular transcription factor that we have termed YY-1.

We first identified the binding site for YY-1 in the P5 transcriptional control region of adeno-associated virus, a defective virus that depends on a variety of adenovirus gene products for its replication. The E1A protein activates expression of the P5 control region, and the critical sequence element required for activation is a DNA segment constituting the binding site for YY-1. To investigate its function, this site was inserted upstream of several heterologous promoters, and it repressed their activity. The repression was relieved in the presence of E1A protein. In fact, the protein also activated transcription through the YY-1 binding site. The combination of these two effects, relief of repression and activation, induced transcription by a factor of 1,000 in some test genes. Thus the combination of the YY-1 binding site and E1A protein formed a powerful biological on/off switch.

In order to study the YY-1 factor, we prepared some from cultured human cells and determined a short amino acid sequence from the purified protein. This sequence was used to design a short probe DNA, which enabled us to identify and isolate a cDNA clone encoding the protein. Sequence analysis of the clone revealed that YY-1 is a 414–amino acid protein with a zinc finger DNA-binding motif. Protein was expressed from the clone and shown to bind specifically to the YY-1 recognition site.

The binding activity of YY-1 was altered by fusing it to the DNA-binding domain of the yeast GAL4 protein. This approach is widely used to study transcription factors, since it provides the opportunity to direct binding of the factor under study in a highly specific fashion to a test gene construct that contains the yeast GAL4 DNA-binding site. By redirecting the factor to bind to a novel site, it was possible to study the activity of the fusion protein in cells that contain high levels of endogenous YY-1. As anticipated, the YY-1 fusion protein repressed transcription of the test gene in the absence of the E1A protein, and the repression was relieved by the E1A protein. Mutational analysis of the fusion protein has demonstrated that the YY-1 zinc finger domain is responsible for its ability to repress transcription.

Since the E1A protein would not be expected to cause the YY-1 fusion protein to detach from the GAL4 DNA-binding site, it appears likely that YY-1 remains bound to the control region but is somehow altered in the presence of the E1A protein so that transcription is not repressed. Several experiments indicate that the E1A protein can bind directly to YY-1. For example, if the two proteins are mixed, they sediment as a complex in a sucrose gradient. Furthermore, radioactively labeled E1A can bind to YY-1 that has been separated from a complex mixture of proteins by electrophoresis and bound to a membrane filter.

We are currently mapping the domains on E1A and YY-1 through which they interact. The location of these domains might provide insight to the mechanism by which the viral activator alters YY-1 function. We are also searching for cellular proteins that might interact with the domain of YY-1 responsible for repression.

The P5 promoter of the adeno-associated virus contains two binding sites for YY-1. The site that was originally studied is centered about 60 base pairs upstream of the transcriptional initiation site, and it is responsible for the repression and E1A-mediated activation discussed above. The second site is centered at the P5 transcriptional control region. Its location suggested that YY-1 might serve an initiator function in addition to its repression function. Short DNA sequences surrounding several transcriptional start sites have been shown to be capable of starting transcription and have been termed initiator elements. These sequences can direct RNA polymerase to initiate transcription at the correct start site in the absence of any other binding sites for known transcription factors. When binding sites for additional factors are added, the initiator sequences become much more efficient.

The binding site for YY-1 proved to behave as an initiator element. Furthermore, by studying its activity in cell-free extracts and using antibodies specific for YY-1, we were able to show that YY-1 protein is required for the initiator activity displayed by its binding site. Work in progress to determine whether YY-1 interacts with other proteins in the initiation complex should elucidate its role in transcription.

It is intriguing that YY-1 (Yin and Yang factor 1) is able to exert opposite effects when it binds to different locations within a promoter. In the adeno-associated virus promoter, it represses transcription when bound upstream of the start site, and it contributes in a positive sense to the initiation event when bound at the start site. Our longer term goal will be to understand how the factor can mediate these two different activities.

Finally, it is important to note that in addition to activating transcription through YY-1 and other cellular transcription factors, the E1A protein can oncogenically transform cells. Almost certainly, at least part of its oncogenic activity results from its ability to bind to proteins such as the cellular retinoblastoma protein, a recessive oncogene product whose proper function is required for appropriate growth regulation of cells. It is possible, however, that interactions of E1A protein with transcriptional regulatory proteins also contribute to oncogenesis. Work is in progress to determine whether alteration of YY-1 function can play a role in cellular transformation.

Growth Control of Myeloid Cells

Charles J. Sherr, M.D., Ph.D.—Investigator

Dr. Sherr is also a member of the Department of Tumor Cell Biology at St. Jude Children's Research Hospital and Adjunct Professor of Biochemistry at the University of Tennessee College of Medicine, Memphis. He received his medical degree and his Ph.D. degree in immunology from New York University School of Medicine, where he studied with Jonathan Uhr. After a pathology residency at Bellevue Hospital Center, New York, he joined George Todaro's laboratory at the National Cancer Institute, where he began studies on retroviral oncogenes. After 10 years on the staff of the NCI, Dr. Sherr relocated to St. Jude Children's Research Hospital.

EACH day humans produce billions of blood cells, which enter the circulation from their sites of origin in the bone marrow. The majority are red cells (erythrocytes), which transport oxygen, and the remainder are white cells (leukocytes), which play a vital role in preventing infection by bacteria, viruses, and other parasites. Different classes of white cells carry out specialized functions: macrophages and granulocytes ingest and kill microorganisms, and lymphocytes recognize foreign antigens and produce antibodies to combat them.

The process of blood cell production (hematopoiesis) is regulated by a group of protein growth factors, termed colony-stimulating factors (CSFs) or interleukins. These factors stimulate the precursors of mature white cells to form colonies in agar composed of differentiated blood cell elements and were named for the types of colonies they produced. For example, M-CSF (or CSF-1) specifically induces macrophage colonies, G-CSF promotes granulocyte development, and GM-CSF stimulates the growth and differentiation of both types of cells. CSFs, now produced in quantity through genetic engineering techniques, have become part of the clinical armamentarium and are efficacious in extrinsically regulating blood cell production and in heightening host defense against infection.

Signal Transduction by the CSF-1 Receptor

The actions of CSFs in supporting cell proliferation and survival are mediated through their binding to specific receptors expressed on the surfaces of their target cells. The macrophage CSF-1 receptor (CSF-1R) consists of an extracellular growth factor–binding portion, joined through a single membrane-spanning segment to an intracellular enzymatic domain. Binding of CSF-1 to its receptor on the outside of the cell triggers the activity of the intracellular enzymatic moiety—a kinase—inducing it to add phosphate molecules to other proteins. These phosphorylation events modify the biochemical behavior of multiple target proteins, some of which relay signals to the cell nucleus that alter gene expression, DNA synthesis, and cell division.

The ability of many cell types to respond to CSF-1 is simply determined by whether they express its receptor. By introducing the gene encoding CSF-1R into naive cells, we can sensitize them to the stimulatory effects of the growth factor. Such manipulations have allowed us to study genetically engineered receptor variants for their capacity to transduce signals for cell growth or differentiation.

One strategy is to alter different portions of the intracellular domain of CSF-1R in order to pinpoint structural motifs that determine its interaction with its "downstream" targets. A consequence has been the development of receptors that are impaired in transducing signals through certain pathways but not others. In particular instances the introduction of other complementing genes into cells that express partially defective receptor mutants has reconstituted full receptor activity, thereby providing genetic evidence for functional relationships between different gene products in signaling pathways.

Such experiments have revealed that the combinatorial actions of target proteins that bind to, or are phosphorylated by, CSF-1R can in part determine the specificity of the biological response in different cell types, thereby influencing decisions governing cell proliferation, survival, and fate.

Oncogenic Potential of CSF-1R

CSF-1R is encoded by the *FMS* proto-oncogene and can be converted by mutations to an "oncoprotein" capable of inducing tumors. Certain mutations in the extracellular domain of CSF-1R can mimic the action of CSF-1 and activate the receptor kinase in the complete absence of the growth factor. The sustained and unregulated signals for cell growth that arise from this class of mutant receptors might naturally contribute to malignancies involving macrophages or their bone marrow progenitor cells.

We have recently used a prospective genetic

approach to identify sites within CSF-1R that, when mutated, can endow the receptor with oncogenic activity. By randomly mutagenizing segments of *FMS* and screening "libraries" of mutated genes for their ability to induce cell transformation, we identified several sites in the receptor where "activating mutations" occur. With such information in hand, it is now possible to search for the presence of similar genetic lesions in the *FMS* genes of myeloid leukemia cells and so determine whether *FMS* mutations play an etiologic role in such diseases.

Role of CSF-1 in Cell Cycle Progression

After proliferating macrophages complete cell division (mitosis, or M phase), they enter an 8- to 10-hour gap phase (G_1), during which they prepare to replicate their chromosomal DNA. The ensuing period of DNA synthesis (S phase) lasts for seven hours, and once DNA replication is complete, the cells enter a second, shorter gap phase (G_2) before dividing again and redistributing copies of duplicated chromosomes to each daughter cell. CSF-1 is only required throughout G_1 for cells to enter S phase, and once DNA synthesis begins, macrophages can complete cell division in the absence of the growth factor. On the other hand, the requirement for persistent CSF-1R–mediated signals throughout the entire G_1 interval implies that the expression of growth factor–responsive genes must be temporally regulated over an 8- to 10-hour period.

Genetic data accumulated through studies of yeasts indicate that cell division–cycle genes called G_1 cyclins act to prepare cells for DNA synthesis. We recently isolated a novel class of "D-type" cyclins from mammalian cells, at least two of which are differentially regulated by CSF-1 during the G_1 interval of the macrophage cell cycle. Related genes are expressed in other cell lineages, where their expression is governed by different growth factors. Our idea is that these cyclins control progression through the G_1 interval in mammalian cells by mechanistically linking early steps in growth factor–mediated signal transduction with the timing of the cell cycle clock. As might be expected, perturbations in the regulation of these cyclins occur in specific types of tumor cells and thus appear to contribute to malignancy.

The Role of Second Messengers in Ion Channel Regulation

Steven A. Siegelbaum, Ph.D.—Associate Investigator

Dr. Siegelbaum is also Associate Professor of Pharmacology in the Center for Neurobiology and Behavior, Columbia University College of Physicians and Surgeons. He received his A.B. degree in biochemical sciences from Harvard College and his Ph.D. degree in pharmacology from Yale University, studying the role of calcium in cardiac electrical activity. He then did postdoctoral research with David Colquhoun at University College, London, and with Philippe Ascher at the Ecole Normale Supérieure in Paris, where he studied the nicotinic acetylcholine receptor ion channel, before joining the faculty of Columbia University. He has received the Herbert J. Kayden Award in Biomedical Science of the New York Academy of Sciences.

THE electrical activity of nerve and muscle cells is regulated by the actions of hormones, neurotransmitters, and sensory stimuli such as light, odors, and pressure. Regulation of neuronal activity often depends on the production of intracellular second messengers—small metabolites such as cyclic AMP, cyclic GMP, and various products of phospholipid metabolism. These second messengers then act to alter the function of ion channels, the membrane proteins that govern the electrical signaling of cells. Previous research in our laboratory focused on the role that regulation of ion channel function by second messengers plays in learning and memory. Recently we have become interested in the role of second messengers in olfactory signal transduction, the subject summarized below.

The second messenger that was first found to play a role in regulating electrical activity was cyclic AMP. From initial studies by Earl Sutherland and his colleagues, a number of hormones and neurotransmitters have been shown to act by elevating cAMP concentrations in cells. Later studies by Edwin Krebs and his colleagues showed that most of the effects of cAMP were due to the activation of a cAMP-dependent protein kinase (cAMP-PK), which phosphorylates many types of proteins. Over the past several years it has become clear that neurotransmitters can alter the activity of ion channels by causing the production of cAMP, leading to the activation of cAMP-PK, which can then directly phosphorylate ion channels. In general, cAMP-dependent actions are relatively slow; they require several seconds to produce changes in electrical activity because of the relatively slow rates of protein phosphorylation and dephosphorylation. Thus phosphorylation-dependent second messenger actions are generally not well suited to mediating rapid neuronal signaling that occurs during fast synaptic transmission or sensory processing.

Role of Second Messengers in Olfactory Signal Transduction

Olfactory signal transduction has provided neurobiologists with an intriguing puzzle on several levels. First, how does the olfactory system recognize and discriminate among thousands of different odors? Are there a limited number of receptors that each bind many hundreds of odorants, or are there hundreds of receptors that are each specific for a single or a few different odorants? Second, how does the binding of an odorant to its receptor generate an electrical signal in the olfactory neuron? Third, how does our olfactory system enable us initially to detect odors at very low concentrations and yet become insensitive to the same stimuli after several minutes? In one sense this form of adaptation is the simplest form of learning: our olfactory system "learns" to ignore a certain stimulus.

Recently some of the puzzles associated with olfaction began to be solved. Linda Buck and Richard Axel (HHMI, Columbia University College of Physicians and Surgeons) identified a surprisingly large gene family that may code for hundreds of distinct odorant receptors in rat olfactory neurons, providing for the requisite specificity. The discrepancy between the slow time course of most cAMP-mediated responses and the more rapid olfactory signaling was resolved when Tadashi Nakamura and Geoffrey Gold reported an ion channel in toad olfactory neurons that was directly activated by cAMP. Thus the activation or gating of the channel did not depend on the relatively slow processes of phosphorylation and dephosphorylation, but rather seemed to be due to the direct binding of cAMP to the channel. This channel was activated equally well by cAMP or cGMP.

Properties of a Cyclic Nucleotide–gated Channel

The first gene for a cyclic nucleotide–gated (CNG) channel to be cloned came from photoreceptors, where the channel participates in phototransduction and is selectively activated by cGMP. More recently several laboratories have cloned the genes for olfactory neuron CNG channels from several species. In collaboration with

Richard Axel's group, we have been studying the properties of a CNG channel cloned from olfactory neurons of catfish. Our goal is to understand how the binding of cyclic nucleotides leads to channel activation and how the channel itself may participate in adaptation.

The gene for the catfish olfactory channel is highly homologous to olfactory channel genes from other species and to the gene for the photoreceptor channel. The channel is expressed selectively in olfactory neurons, supporting its role in olfactory signal transduction. Like the rat olfactory CNG channel, the catfish channel is also activated directly by both cAMP and cGMP. Unlike the rat channel, however, the catfish channel does not discriminate between cAMP and cGMP, suggesting a structural difference in the cyclic nucleotide–binding sites. Using genetic engineering, we are trying to define the structural bases for these differences.

We have also used single-channel recording to measure the unitary currents that flow through an open CNG channel. We find that the probability of a channel being open increases with rising cyclic nucleotide concentration. When the cloned channel is expressed in frog oocytes, it requires relatively high concentrations of cAMP or cGMP (around 50 micromolar) to become activated. Surprisingly, the same channel studied in its native environment, the catfish olfactory neuron, requires 20-fold lower cyclic nucleotide concentrations for activation. This difference in cyclic nucleotide sensitivity could mean that the oocytes fail to process the channel correctly (for example, by not phosphorylating it properly) or that they lack some important channel-regulating protein. Defining the factors responsible for the discrepancy between the cloned and native channels is an important goal.

In other work, we are focusing on the role of the channel in odor adaptation. When the olfactory CNG channel is activated, it allows calcium ions to enter the olfactory neuron from the outside medium. In photoreceptors, a similar influx of calcium underlies visual adaptation. We find that intracellular calcium greatly reduces the response of the olfactory channel to cyclic nucleotides. This effect appears to result from a shift in the dose-response curve for channel activation to higher concentrations of cyclic nucleotides. The inhibitory effect of calcium occurs at physiological calcium levels. Moreover, the effect is not due to a direct action of calcium on the channel but rather appears to involve an intermediate regulatory protein that is loosely associated with the channel. Thus calcium acts as a negative feedback regulator of olfactory responses.

Thus the olfactory system provides a useful model for studying neuronal signal transduction and neuronal plasticity. Studies on the molecular bases of these phenomena should provide us with insight into many of the basic mechanisms controlling nerve cell behavior.

Chemistry of Cellular Regulation

Paul B. Sigler, M.D., Ph.D.—Investigator

Dr. Sigler is also Professor of Molecular Biophysics and Biochemistry at Yale University. He studied chemistry at Princeton University and received his M.D. degree from Columbia University. He then spent two years as a house officer in the Department of Medicine at Columbia-Presbyterian Medical Center, New York. He began his work on crystallography with David Davies at NIH. He studied as a Helen Hay Whitney Fellow at the MRC Laboratory of Molecular Biology in Cambridge, England, where he received his Ph.D. degree in biochemistry. Before accepting his present position, Dr. Sigler was Professor of Biochemistry and Molecular Biology at the University of Chicago. Dr. Sigler was recently elected to the National Academy of Sciences.

MY colleagues and I continue to pursue our interest in the molecular mechanism of regulation at the level of transcription and transmembrane signaling. Our approach is to use high-resolution x-ray crystallography to determine the structure of the relevant macromolecules and their assemblies. We infer chemical mechanisms from these structures and test them with biochemical and physicochemical experiments and with directed mutagenesis.

Transcriptional Regulation

In the area of transcriptional control, our primary focus remains the mechanism by which DNA-binding regulatory proteins are targeted to their responsive genes. Our earliest and best-studied experimental system has been the allosteric regulation of the *trp* repressor and the formation of the *trp* repressor-operator complex. Having defined the stereochemistry of the protein-DNA interface, we are now extending the study (partly in collaboration with Julian Sturtevant, Yale University) to include an analysis of the thermodynamic parameters underlying the *trp* repressor's ability to bind selectively to the *trp* operator.

Crystallographic structural analyses have been extended to the mechanism by which steroid receptors bind selectively to their DNA targets. The first complex in this series (in collaboration with Len Freedman and Keith Yamamoto, University of California, San Francisco) shows the DNA-binding domain in complex with its idealized target, the symmetrical glucocorticoid response element. It shows how the two zinc fingers of the domain interact with each other to form a unified globular domain and how the domain makes chemical contacts with the bases that identify the "half sites" of the symmetrical response element. It also shows a surprisingly novel mechanism by which the receptor recognizes the invariant three–base pair spacing between the half sites.

Interactions with the DNA phosphates stabilize a change in the protein's structure that causes it to form a firm dimer. In so doing, the DNA-binding surfaces of each subunit are placed in perfect registry with the complementary surfaces of the DNA target's half sites. Thus the DNA itself helps form a dimer that discriminates between DNA sequences that have the correct spacings between their half sites and those that do not. These same DNA-induced conformational changes are also likely to potentiate functional contacts between the receptor and other components of the transcription initiation assembly.

Work is under way to examine the arrangement by which the glucocorticoid receptor contacts other transcription factors on "composite" regulatory sites. We are also examining the way other members of this steroid receptor superfamily interact with their respective response elements. Only through a comparative study can we build a reliable picture of the targeting mechanism.

Most recently we have solved (in collaboration with Laimonis Laimins, HHMI, University of Chicago) the structure of the complex between the E2 transcription factor of bovine papilloma virus and the DNA sequence (enhancer) to which it binds. E2 is involved in replication of the viral genome and controls expression of the genes responsible for transformation of cells infected with this oncogenic virus. The structure has been refined at an unprecedented level of detail (1.7 Å) and exhibits a barrel-like dimeric architecture that has never been seen before in any protein structure, let alone in a DNA-binding domain. The DNA is severely bent over this barrel as it forms specific contacts with the protein. These results are now being correlated with genetic information produced by others.

Several other specific DNA complexes of transcriptional regulatory proteins are in earlier stages of crystallographic analysis. A particularly interesting one is the *arg* repressor, which regulates the expression of arginine biosynthetic genes. It is unique because it is a hexamer of identical subunits. Its target is also unusual, in that it is two symmetrical 18–base pair sequences whose symmetry axes are separated by 21 base

pairs, or exactly two helical turns of DNA. Our studies (in collaboration with Werner Maas, New York University) support the inference that the operator must be sharply bent as it wraps around the hexamer. In view of the distortion that this protein can introduce into DNA, it is interesting that the *arg* repressor has been implicated as a binding "scaffold" in the site-specific recombination process that resolves fused ColE1 plasmids. The architecture of this protein-DNA complex should give insight into the DNA topology involved in site-specific recombination.

Transmembrane Signaling

In the area of transmembrane signaling, we have recently solved (in collaboration with Jeff Browning of Biogen) the structure of the human secretory phospholipase A_2 that is associated with both acute and chronic inflammation. This study contrasts the free enzyme with the enzyme in a complex with a transition-state analogue and shows that the mechanism of this enzyme conforms to what we have proposed earlier for interfacial catalysis. This structure could provide a firm stereochemical platform for anti-inflammatory drug design.

In order to function, this class of enzymes must attach itself firmly to the surface of the membrane. Others have shown that the attachment of the enzyme to the membrane surface is primarily due to electrostatic forces. Our calculations (in collaboration with Barry Honig, Columbia University) of the charge-potential distribution for a large series of phospholipase A_2 structures account nicely for this effect.

Recently our attention has turned to the transmembrane signal cascade that employs seven-helical transmembrane receptors, G proteins, and target enzymes involved in either the synthesis or cleavage of phosphodiester bonds (e.g., adenylcyclase, phospholipase C, cGMP phosphodiesterase). Because of the immediate availability of large amounts of pure protein, we have chosen to work first on the bovine visual receptor system. Our first goal is to solve a recently obtained well-ordered crystal form of $T_{G\alpha} \cdot GTP$, the soluble transducer of the activation signal. We hope to extend this work to include complexes of $T_{G\alpha}$ with the activated receptor (rhodopsin) and its target molecule, the γ-subunit cGMP phosphodiesterase.

The Mitochondrial Genome of Trypanosomes

Larry Simpson, Ph.D.—Investigator
Dr. Simpson is also Professor of Cell and Molecular Biology at the University of California, Los Angeles. He received his B.A. degree in biology from Princeton University and his Ph.D. degree in molecular parasitology with William Trager at the Rockefeller University. His postdoctoral training was with Maurice Steinert at the Free University of Bruxelles.

THE kinetoplastids represent a large group of parasitic protozoa that are the causal agents for a variety of human and animal diseases, including African sleeping sickness, Chagas' disease, and leishmaniasis. There are no vaccines for any of these, and chemotherapies are either nonexistent or inadequate. As cells, the kinetoplastids represent one of the most ancient lineages of the eukaryotic kingdom and thereby have many novel physical and biochemical features. Some species, such as *Crithidia*, are parasitic in only a single invertebrate host. Others, such as *Leishmania*, *Trypanosoma*, and *Phytomonas*, are parasitic in both an invertebrate and a vertebrate (or plant) host.

Kinetoplast DNA

The mitochondrial genome of these cells, which is known as kinetoplast DNA, consists of a single giant network containing approximately 10,000 minicircles and 20–50 maxicircles, all linked together by catenation. The maxicircles comprise a subset of mitochondrial genes that encodes two small rRNAs, three subunits of cytochrome oxidase, cytochrome b, four subunits of NADH dehydrogenase, and three yet unidentified proteins. No tRNAs appear to be encoded by the mitochondrial genome and therefore must be imported into the organelle by some yet to be defined mechanism. The function of the minicircle was unknown until recently.

RNA Editing

RNA editing is a post-transcriptional process in which uridine (U) residues are inserted and deleted from coding regions of the primary transcripts of several maxicircle "cryptogenes." The extent of editing varies from a few U residues to hundreds, at hundreds of sites throughout the entire mRNA (pan-editing). The function of editing is to create translatable mRNAs encoding mitochondrial proteins.

The extent of editing of specific genes varies from species to species. For example, the NADH dehydrogenase subunit 7 mRNA is internal and 5′-edited in *Leishmania tarentolae* but pan-edited in *Trypanosoma brucei*. We have recently shown that at least one G-rich intergenic region in both species encodes a transcript that is pan-edited to produce ribosomal protein S12 for the mitochondrial ribosome. It is likely that five other G-rich regions represent additional pan-edited cryptogenes, which would bring the total mitochondrial structural gene content in these cells to 17, of which 12 are cryptogenes.

Mechanism of RNA Editing

We discovered in 1990 that maxicircles also encode another class of RNAs—the guide RNAs (gRNAs)—which contain the necessary sequence information for the edited genes. These are small RNAs, which can form perfect duplex hybrids with edited mRNAs, both within the edited region and 3′ of the edited region, provided G-U base pairs are allowed. The gRNAs also have nonencoded 3′ oligo-[U] tails ranging in length from 5 to 28 nucleotides. We then discovered that gRNAs are also encoded in the minicircles, finally providing a genetic role for these enigmatic molecules.

We have proposed two models for the involvement of gRNAs in editing. In both models the initial interaction of the gRNA and the mRNA is the formation of an anchor duplex just downstream of the pre-edited region. The "enzyme cascade" model invokes an endonuclease that cleaves at the first mismatch, a terminal uridylyl transferase that adds a U to the 3′-hydroxyl, and an RNA ligase joining the two mRNA fragments. In the "transesterification" model, the added U's are derived from the oligo-[U] tail of the gRNA by means of two successive transesterifications. In both models the gRNA provides an internal guide sequence to specify the precise addition and deletion of U's at specific sites.

A major goal is to obtain an *in vitro* system in which the entire editing process occurs in an accurate manner, thereby allowing a biochemical dissection and reconstitution of the underlying enzymatic machinery. We have recently shown that at least the initial step of RNA editing—the formation of gRNA-mRNA chimeric molecules—

can be performed *in vitro* by incubation of synthetic RNAs with a mitochondrial extract and that this process is anchor-dependent.

The work on RNA editing was supported by a grant from the National Institutes of Health.

Evolutionary Considerations

The existence of split genes is not novel, but the existence of genes in which the RNA product of one gene contains information for the correction of coding sequences within transcripts of the other gene is novel. If the transesterification model proves to be correct, this would suggest that RNA editing is on the same evolutionary pathway as RNA splicing and may in fact represent a primitive type of trans-splicing with partial integration. The trypanosome type of RNA editing has not yet been reported in other organisms, but other types of modifications of the sequences of coding RNAs have been observed in organisms as diverse as plants and humans. C to U changes at specific sites occur in transcripts of several human genes and also in transcripts from many plant mitochondrial and chloroplast genes. The determination of site specificity for these multiple transitions is completely unknown.

The evolution of RNA editing in the trypanosomes is interesting in itself. To investigate it, we plan to examine representatives of more-primitive kinetoplastid lineages.

The Kinetoplast Genome as a Target for Disease Diagnosis

We have shown that the kinetoplast DNA minicircle molecule of *Trypanosoma cruzi*, the causal agent of Chagas' disease, is an appropriate multicopy target for detection and strain classification of the parasite in patients, animals, or insects. The minicircle of *T. cruzi* consists of four conserved regions and four variable regions. There are multiple minicircle sequence classes in a network, and minicircles in different strains are very polymorphic. Primers to the conserved region were used for polymerase chain reaction (PCR) amplification of minicircle fragments from either the conserved region or the variable region. This method is being developed into a diagnostic procedure to detect small numbers of parasites in blood of chronically ill patients and to classify the strain of the parasite.

We recently developed a method to recover blood from patients and preserve total DNA without refrigeration. The DNA is then cleaved with a chemical nuclease to release linearized minicircles, and the fragments are amplified with specific primer sets. We are attempting to expand this diagnostic procedure by developing a multiplex PCR-based assay for the detection of multiple blood-borne viral and parasitic disease agents.

The kinetoplast DNA minicircle has also proved to be an appropriate target for the diagnosis of other pathogenic kinetoplastid infections, such as those caused by *Leishmania* species.

Grants from the World Health Organization and the Rockefeller Foundation supported this diagnostic work.

RNA Editing as Possible Target for Intervention

Whenever a parasite has a biochemical pathway that is unique to the parasite and not found in the human host, there is a potential for selective chemotherapy. As RNA editing is dissected biochemically, we plan to examine the possibility of inhibiting components of the editing machinery in the parasites without affecting the host. This may open up a new direction for selective chemotherapy of the many trypanosome-caused diseases.

Regulation of Gene Activity During B Cell Development

Harinder Singh, Ph.D.—Assistant Investigator

Dr. Singh is also Assistant Professor in the Department of Molecular Genetics and Cell Biology at the University of Chicago. He received his Ph.D. degree in biochemistry, molecular biology, and cell biology with Lawrence Dumas at Northwestern University. His postdoctoral research was done with Phillip Sharp at the Massachusetts Institute of Technology, as a Jane Coffin Childs fellow. He remained at MIT as a research associate until his move to Chicago.

MY research interests are focused on the analysis of transcriptional regulatory circuits that turn genes on or off during the growth and differentiation of B lymphocytes, cells that produce antibodies. The B cell lineage is a very useful model for exploring the molecular basis of differential gene activity in mammalian development. We are seeking answers to the following questions. What is the nature of the genetic element(s) linked to a target gene that controls its transcriptional activity? What is the nature of the regulatory protein(s) that recognizes this genetic element? How upon binding near its target gene does the regulatory protein modulate the activity of the enzyme complex that transcribes the gene? What is the mechanism by which a growth or developmental signal is transduced by the regulatory protein(s) to effect gene activity? How does the structure of the chromatin within which the target gene is packaged influence the function of the regulatory protein(s)?

The heavy- and light-chain genes encoding the immunoglobulin (Ig) molecule are selectively transcribed in B cells. These genes are assembled from gene segments through an ordered series of somatic recombination events that occur in a developing B cell. The heavy-chain gene locus is the first to undergo recombination and transcriptional activation, thereby defining the pre–B cell developmental state. Subsequently one of two light-chain gene loci is recombined and expressed, resulting in the development of a mature B cell. Ig genes contain multiple cis-acting transcriptional regulatory elements that restrict their expression to appropriate stages in the B lineage.

Previous work by various research groups has resulted in the identification, characterization, and cloning of a B cell–specific regulatory protein, Oct-2. This protein recognizes the octanucleotide sequence ATTTGCAT, which confers B cell specificity to Ig gene promoters. The same sequence motif is also a functional component of the heavy-chain and κ-light-chain gene enhancers. Oct-2 can activate transcription of a reporter gene linked to an Ig promoter in a non–B cell. Thus Oct-2 appears to be both necessary and sufficient for regulating the activity of Ig promoters. The deduced amino acid sequence of Oct-2 reveals a region of similarity that is shared with three other regulatory proteins: Pit-1, Oct-1, and unc-86. This region is termed the POU box and includes a subdomain related to the homeobox.

The levels of the Oct-2 protein are regulated during B cell differentiation. The Oct-2 protein is expressed at low levels in pre–B cells. Oct-2 levels are elevated 5- to 10-fold upon pre–B cell differentiation, and higher Oct-2 protein levels correlate with activation of the Ig κ-light-chain gene locus. Increased expression of Oct-2 in pre–B cells is induced by signaling with the B cell mitogen, bacterial lipopolysaccharide, as well as the lymphokine interleukin-1 (IL-1). Transforming growth factor-β, an inhibitor of κ gene induction in pre–B cells, blocks the up-regulation of Oct-2 but not the activation of NF-κB. The latter is another regulatory protein that has been previously implicated in controlling the activity of the κ locus. We have proposed a model in which the concerted action of increased levels of Oct-2 and activated NF-κB controls the proper stage-specific expression of the κ locus. To test genetically the function of Oct-2 in regulating immunoglobulin gene transcription, we have used gene targeting to disrupt sequentially the two copies of the Oct-2 gene in a B cell. The analysis of the mutant and wild-type cells is currently under way.

Expression of the Oct-2 gene has been shown to be regulated at the level of transcription during pre–B cell differentiation. By studying the control of Oct-2 expression, regulatory proteins interacting with Oct-2 in the genetic hierarchy underlying B cell differentiation can be identified and isolated. A promoter controlling transcription of the Oct-2 gene has been identified. This region is being analyzed and will enable the characterization of transcription factors that regulate Oct-2 expression as well as B cell development.

B cells need to activate the expression of two other genes, *mb-1* and *B29*, to function as antigen-recognizing cells. These genes encode membrane proteins that associate with the antibody

molecule (antigen receptor) and are required for the expression of the receptor on the cell surface. The *mb-1* gene is activated early in B cell ontogeny, continues to be expressed in mature B cells, but is turned off in plasma cells. We have identified the *mb-1* promoter and shown that it contains a regulatory domain that functions in a cell-type and stage-specific manner. A 25–base pair element within this domain is necessary and sufficient for activity. This element is recognized by a novel transcriptional activator termed BLyF, whose distribution correlates with the pattern of expression of *mb-1* in B lineage cells. Thus BLyF may play an important role in B cell development, in part by regulating the activity of the *mb-1* gene.

Regulation of Gene Expression in Developing Lymphocytes

Stephen T. Smale, Ph.D.—Assistant Investigator

Dr. Smale is also Assistant Professor of Microbiology and Immunology and a member of the Molecular Biology Institute at the University of California School of Medicine, Los Angeles. He received his Ph.D. degree in biochemistry from the University of California, Berkeley, where he studied with Robert Tjian. Dr. Smale's postdoctoral research was done with David Baltimore at the Whitehead Institute, Massachusetts Institute of Technology.

HEMATOPOIESIS refers to a complex developmental process through which pluripotent stem cells in mammalian fetal liver or adult bone marrow give rise to several types of terminally differentiated blood cells. The immune system incorporates many of these cell types, which include B and T lymphocytes, granulocytes, monocytes, and killer cells, to protect the host from infection by a variety of means. When defects arise in the hematopoietic pathway, the effects are often severe. In some cases, precursor cells cannot develop to maturity, resulting in immunodeficiency. In other cases, uncontrolled proliferation of developing cells results in leukemias and lymphomas. To understand the basis of these defects, we must explain the regulation of hematopoiesis at the molecular level.

The approach used in our laboratory to study the regulation of the hematopoietic pathway is to identify proteins that directly activate or inactivate genes expressed at specific stages of B and T lymphocyte development. Lymphocytes play a central role in the immune response by mediating the recognition of foreign and infectious matter. As B and T cells mature, a wide variety of genes are turned on and off in a specific temporal pattern. An analysis of the mechanisms by which these genes are precisely controlled is a good starting point for studying the regulation of hematopoiesis.

Our laboratory currently focuses on the regulation of the gene encoding terminal deoxynucleotidyltransferase (TdT). This gene is turned on for only a short time during both B and T cell development. The TdT protein is a template-independent DNA polymerase that appears to play a role in generating diversity within the antibody and T cell receptor molecules.

We chose to analyze the TdT gene because of its expression patterns in normal and leukemic cells. In normal cells TdT expression is unusual because it is found in both early B and T cells, suggesting that it may be regulated by transcription factors common to these two related but distinct lineages. Moreover, TdT is expressed at high levels in acute lymphocytic leukemias (ALLs) and at lower levels in many acute myeloid leukemias (AMLs). TdT expression in AMLs is particularly intriguing because it does not occur in normal myeloid cells or myeloid precursors. Therefore the abnormal expression of TdT in AMLs suggests that the protein or proteins that deregulate TdT may also play a role in leukemogenesis.

To study the mechanisms of regulating TdT expression, we have analyzed the promoter, the transcriptional control region surrounding the start site for TdT RNA synthesis. Our analysis focuses on two aspects of promoter function: 1) the DNA sequence elements and proteins responsible for directing TdT expression specifically in lymphoid cells and 2) the unusual architecture of the TdT promoter region, which is fundamentally different from that of most other mammalian promoters that have been studied in detail. Our progress in these two areas is described below.

Regulated Expression of the TdT Gene

To understand the specific activation of the TdT gene in B and T lymphocytes, we are currently focusing on a DNA sequence element located 60 nucleotides upstream from the TdT transcription start site. We have shown that this element of approximately 15 base pairs, called D′, is essential for efficient promoter activity in lymphocytes. We have identified two different classes of proteins that interact with the D′ region, one or both of which may be important for TdT transcription. We identified the first protein, called LyF-1, by looking for protein-DNA interactions at the D′ site in protein extracts derived from lymphocytes. We have characterized this novel protein after purifying it from a murine thymoma cell line. LyF-1 is a 50-kDa protein and was found to be highly enriched in lymphoid cells relative to a variety of nonlymphoid cells. Moreover, LyF-1 was found to bind to two different sites in the TdT promoter as well as to the promoters for four other lymphocyte-specific genes.

The second class of proteins that bind to the TdT D′ site are those related to the ets family of DNA-binding proteins. Mammalian ets-1 was first

identified as the cellular homologue to a retroviral oncogene and was shown only recently to bind to specific DNA sequence elements. As with most genes expressing DNA-binding proteins, the *ets-1* gene is related to several other genes found in the mammalian genome, all of which encode proteins that bind to similar DNA sequence elements. We have expressed in bacteria three different members of the ets family—ets-1, fli-1, and PU.1—and have found that all three bind tightly to the TdT D′ site. Interestingly, all three of these proteins are expressed at high levels in certain subsets of hematopoietic cells, and both ets-1 and fli-1 are expressed in most, if not all, cell types that express the TdT gene.

We have determined that LyF-1 is not ets-1 or fli-1 and is unlikely to be a member of the ets family of proteins. Therefore our current goal is to determine which of these proteins—LyF-1, ets-1, fli-1, or a combination of LyF-1 and an ets member—is required to activate TdT transcription by interacting with the D′ element. Those proteins required for TdT activation are likely to be important for regulating lymphocyte development and will become candidates for involvement in the deregulation of TdT expression in ALL and AML. This work is supported by a grant from the National Institutes of Health.

Unusual Design of the TdT Promoter

The TdT promoter is unusual in that it does not contain a TATA box, which is a common DNA sequence element found in most promoters that have been characterized in detail. The TATA element is known to carry out several important functions in promoters, including 1) recruitment of the RNA polymerase to the transcription start site region, 2) determination of the site's location, and 3) determination of the direction of transcription from the promoter. To understand the regulation of the TdT promoter, we are inquiring into how these functions are carried out in the absence of a TATA box. This question is especially important for an understanding of transcriptional control during lymphocyte development, since it appears that the promoters of most genes expressed during this process do not contain TATA boxes.

Previously we found that in place of the TATA element, which is typically located 30 nucleotides upstream of the transcription start site, the TdT promoter contains a distinct element that instead overlaps the start site. This element, which we call an initiator (Inr), is like the TATA element in its importance for promoter function and also in pinpointing the RNA start site to a specific nucleotide.

We are now comparing further the activities and mechanisms of action of TATA boxes and Inr elements. Currently our data suggest that the mechanisms for TATA and Inr-mediated transcription follow very similar steps prior to the actual onset of RNA synthesis.

We would also like to understand the contributions of TATA and Inr elements toward determining the start site location and the direction of transcription within a promoter. Some promoters contain both TATA and Inr elements. Although both elements can influence start site localization and transcriptional directionality, the TATA element appears to be much more powerful than an Inr. Our results, however, also have led us to challenge the belief that the orientation of a TATA box within a promoter alone determines the direction of transcription. Instead, our data suggest that the direction of transcription within a simple promoter is determined by the location of a TATA box or Inr element relative to a specific activator element.

We are continuing to define the rules by which a promoter determines the direction of RNA synthesis in order to understand further the design of more-complex promoters, including those that contain, like the TdT promoter, important sequence elements both upstream and downstream of the transcription start site.

Developmental Genetics

Philippe M. Soriano, Ph.D., D.Sc.—Assistant Investigator

Dr. Soriano is also Assistant Professor at the Institute for Molecular Genetics and Department of Cell Biology, Baylor College of Medicine. He obtained his Ph.D. degree in biochemistry and his D.Sc. degree from the University of Paris. He did postdoctoral research in France, and then with Rudolf Jaenisch in Germany and at the Whitehead Institute for Biomedical Research of the Massachusetts Institute of Technology, before joining the faculty at Baylor. Dr. Soriano is a Pew Scholar in the biomedical sciences.

THE major aim of our research is to extend the understanding of early development of the mouse. Our approach is to create mice that are mutant in genes critical for this process. Two techniques are being used: random mutagenesis, which should discover unknown genes, and targeted mutagenesis in previously characterized genes for which no mutation is known.

The random mutagenesis project is designed to identify developmentally regulated and critical genes in the embryo. Identification of the mutated gene in many of the classical mouse mutations has been difficult because there is no convenient tag by which to clone the mutation. Insertion mutagenesis in transgenic mice, wherein a gene is disrupted by introducing a foreign fragment of DNA into the germline, is attractive, since the transgene can serve as a tag. However, only 1 out of 20 transgenic strains displays an overt phenotype, so the approach is laborious and time consuming. We have circumvented this problem by preselecting for mutations in embryonic stem (ES) cells, which are then introduced into embryos to colonize the germline.

The selection procedure we have used is termed "promoter trapping." In this method, a reporter gene is placed downstream of a splice acceptor, and the construct is then introduced into ES cells. Expression of the reporter gene can only originate from a flanking cellular promoter. Therefore transgenic mice derived from ES cells selected for such events can be used both to trace the activity of the tagged gene, by expression of the reporter gene, and to mutate the gene. The reporter gene we have used encodes a fusion protein with both β-galactosidase (β-gal) and neomycin phosphotransferase (neo) activity. This allows one to select directly for promoter trap events, since neo confers resistance to the drug G418, and to follow the activity of the trapped gene by incubating the embryos in a dye that turns cells blue if β-gal is present.

Thirty-one transgenic lines have now been generated using a retroviral promoter trap vector and have been examined both for patterns of expression and for phenotype. We are particularly interested in strains that exhibit restricted patterns of expression at gastrulation, a critical stage of development in the early mouse embryo. Twelve strains carry a recessive lethal mutation due to the promoter trap insertion, and one induces sterility. This demonstrates that the method is very valuable for isolating developmental mutants.

Further analysis of one of the embryonic lethal strains suggests a mutation in a transcription factor. The other strains, which do not display a phenotype associated with the promoter trap event, may reflect mutations in nonessential genes. This first aspect of our research is supported by the National Institutes of Health.

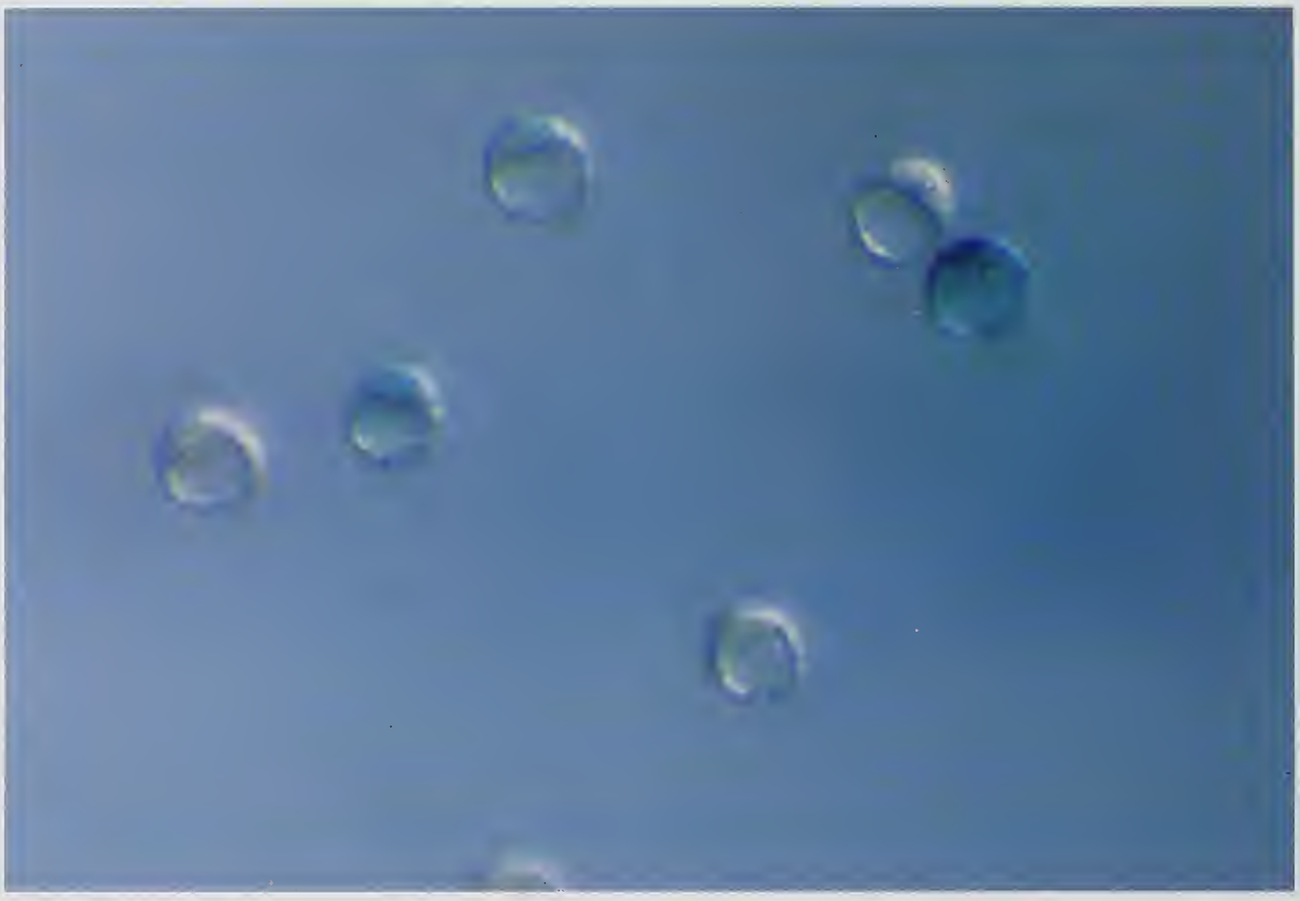
It is also possible to make specific mutations by targeting genes believed to play critical roles in development, based on their previous characterization. This approach relies on the ability to introduce into ES cells mutant gene constructs that will recombine homologously with the normal gene, at the correct chromosomal location, therefore creating a mutation in the gene of interest. Our initial efforts focused on the gene encoding c-*src*, the first proto-oncogene identified.

Src is a protein-tyrosine kinase, broadly expressed but at particularly high levels in neurons and in platelets. The *src* gene was knocked out in ES cells, and the mutation was transmitted through the germline. Surprisingly, animals homozygous for the mutation do not die at an early stage, but develop osteopetrosis, a disease characterized by an impaired function of the osteoclasts, the cells that normally resorb bone. As a result of the mutation, the mutant mice fail to undergo normal tooth eruption and have to be maintained on a soft food diet. We have been able to show that the defect lies in the osteoclasts proper, rather than in accessory cells such as osteoblasts that condition osteoclast activity. However, no defect is found in cells or tissues where *src* is most highly expressed.

To explain the lack of a more severe phenotype, we have tested the hypothesis that another *src*-related kinase may be compensating for loss of *src* in the mutant mice. Mice mutant in the closely related c-*yes* and *fyn* genes have therefore been generated by homologous recombination in ES cells. Neither of these mutations results in an

overt phenotype. However, further analysis of the *fyn* mutation has revealed that thymocytes are unable to proliferate in response to stimulation through the T cell receptor, demonstrating that this kinase is essential for normal thymocyte function. Surprisingly, peripheral T cells can respond through the T cell receptor, implying that other kinase(s) may be involved. This mutation may have further effects on the repertoire of T cells, and further studies are under way.

To test the concept that *src*-related kinases compensate for one another, we have crossbred the mutant mice and examined progeny for double mutant genotypes. The vast majority of *src-yes* and *src-fyn* double mutants die at birth—supporting the compensation hypothesis. The nature of the defect in the double mutants is under further investigation. This second aspect of our research has been supported in part by the National Institutes of Health.



Expression pattern of the β-galactosidase (β-gal) gene in blastocysts of a "promoter trap" strain of transgenic mice, showing labeling (blue) in both the inner cell mass and the outlying trophectoderm. By this technique, mice derived from selected embryonic stem cells can be used both to trace the activity of a β-gal–tagged gene and to mutate the strain.

Research of Glenn Friedrich and Philippe Soriano.

Understanding How Eggs Work

Allan C. Spradling, Ph.D.—Investigator

Dr. Spradling is also a staff member of the Department of Embryology at the Carnegie Institution of Washington in Baltimore and Adjunct Professor of Biology and Microbiology at the Johns Hopkins University. He earned his B.A. degree in physics from the University of Chicago and his Ph.D. in cell biology from the Massachusetts Institute of Technology. His postdoctoral study was done at Indiana University with Anthony Mahowald. Dr. Spradling is a member of the National Academy of Sciences and has received many honors for his work.

ALTHOUGH the union of egg and sperm initiates the complex processes that ultimately result in a new animal, the roles played by the two cells in embryonic development are by no means equivalent. Even nonspecific treatments, such as pricking with a needle, will stimulate many types of eggs to develop in the absence of sperm. Without an egg, however, neither a sperm cell nor any other can even begin the processes that lead to an embryo and ultimately an adult. In fact, eggs have undergone extensive and intricate preparations that allow them to support and direct embryogenesis. If the mystery of a new life could be condensed into a single cell, that cell would surely be an egg.

Eggs Pose Exceptional Biological Problems

Egg cells are very different in structure and in biological capacity from other cells. Most chicken cells weigh less than one millionth of an ounce, whereas a single chicken egg makes a nice breakfast. Eggs not only differ in size but in many other important ways. For example, the egg's genes function differently from those of other cells; so extensive are these differences that the chromosomes appear in the microscope quite unlike those of other tissues. In eggs the products of many genes accumulate and are stored in special forms so that they can be utilized at precisely the right times during embryonic development. Indeed, storage of material is one of the reasons many eggs are enormous.

A great deal remains to be learned about egg structure. How are gene products specially stored, and what are their specific functions later in development? Is each product located in a particular place? What are the mechanisms that allow these materials to be produced and stockpiled in the appropriate manner? The sheer complexity of an egg, with its tens of thousands of specific, highly organized components, has until recently blocked all attempts to describe egg structure in molecular detail, much less to understand the logic that allows this structure to develop into an even more intricate organism. However, over the past 10 years or so, powerful new methods in molecular biology, such as gene cloning, gene transfer, and fruit fly genetics, have begun to unravel the fascinating secrets stored within an eggshell.

Using Genetics to Study Eggs

It is now possible to study structures as complex as eggs because techniques of both genetics and molecular biology allow us to take eggs apart gene by gene. Each component of an egg is specified by a gene carried on the chromosomes. Much research on eggs utilizes the fruit fly *Drosophila*, since far more has been revealed about its genes during 70 years of genetic studies than about those of any other complex organism, including humans. A genetic mutation in the fly inactivates one gene among the 20,000 to 50,000, and if that gene specifies an egg component, then the mutant fly will produce defective eggs.

Frequently, however, the eggs can still partially function. For example, in a particular mutant, development might begin normally, but then stop because of a lack of stored food. By studying what goes wrong with the mutant eggs, biologists can deduce the normal function of a particular gene's product. In this case, one might conclude that the gene was involved in producing the stored food. To understand eggs, it will be necessary, at a minimum, to create mutations and carry out such studies on all the fruit fly genes, one by one.

If researchers were limited to looking at egg defects in a microscope, progress would still be slow. Many of the defective eggs would not contain any detectable problems; they just wouldn't work. It is at this point that the ability to clone the gene molecularly becomes essential to further advancement. Cloning is simply a way of purifying an individual gene so that its DNA structure can be determined in the laboratory.

There are two major benefits. First, the structure of the gene's product, a specific protein that actually becomes part of the egg, can now be determined by the process of DNA sequencing. Second, specific antibodies can be prepared that only bind to the product of that particular gene.

When these antibodies are used, eggs that previously appeared normal in the microscope can suddenly be seen to have specific defects. Such molecular biological studies allow us to understand in much greater detail where in an egg particular gene products are located, where they go as development begins, and what role they are likely to play in the process.

Transporting Materials Into Eggs

One project of our research group concerns how the massive amount of material found in an egg actually gets there. In the fruit fly most of the egg contents are piped in from 15 nearby cells called nurse cells. Four of these cells are attached directly to the egg by special "pipes" made of the same material as the outer surface of cells, and the remaining 11 are attached to other nurse cells by similar pipes. Thus some of the material in an egg has been shipped through as many as four nurse cells on the way to its destination. Nurse cells not only control somehow the rate and direction of transport, but even select certain products for priority delivery, moving them well ahead of the majority of the egg's contents.

To learn more about the equipment flies use to make their eggs, we identified some mutant strains in which either the connections between nurse cells and egg did not form properly, or material did not flow very well. We have now studied one mutant in some detail. When mother flies lack this gene, they are able to attach on average only four nurse cells to their eggs. As a result the eggs usually do not grow properly and are unable to function. Rarely, a developing egg becomes connected to as many as 11 nurse cells, and a few of these eggs can develop, so the affected females are not completely sterile.

We determined the structure of the protein that was missing in the mutant, and were surprised to learn that it was very similar to one found in large amounts just under the surface of human and other mammalian red blood cells. This protein, called "adducin," has been studied for some time by Vann Bennett (HHMI, Duke University). It has been found at various levels in many types of cell, where it is thought to strengthen cell membranes by acting as a kind of skeleton supporting the floppy lipid bilayer. Red cell adducin is thought to act with other membrane skeleton proteins to mold red cells into their characteristic concave disk shape and to lend the physical properties that let them pass through small capillaries. As with many human proteins, however, the actual function is uncertain because the effects of removing a red cell's adducin are unknown.

At present we are attempting to learn in detail what adducin does to help nurse cells and eggs hook up properly. The pipes form during cell division. Normal daughter cells separate when a region of membrane clamps down as though constricted by a knot from outside. Eventually the two cells are choked off completely. During the last few divisions that give rise to the nurse cells and egg, these constrictions simply stop short, leaving a small interconnecting region of cell membrane. This junction subsequently forms into a pipe. In the adducin mutants, many of the constrictions are unable to stop, so cells that should have built a connecting pipe become completely separated. The pipes that do form in the mutant are much weaker looking than normal, may not transport materials very well, and appear susceptible to breakage.

Understanding how adducin contributes to fruit fly egg production may provide insight into the details of how it works in humans. The mutant flies also have several other abnormalities besides their inability to connect nurse cells and the egg properly, indicating that the function of adducin can be studied in a variety of different situations. Experimenters can now make whatever changes in the protein's structure they wish, introduce a modified gene encoding the altered protein back into the mutant strain of fruit fly, and see if the modified protein will "rescue" some or all of the fly's problems or create new ones. This approach can be applied to the many different gene products that make up a fruit fly egg.

Structural Studies of Regulatory Proteins

Stephen R. Sprang, Ph.D.—Associate Investigator

Dr. Sprang is also Associate Professor of Biochemistry at the University of Texas Southwestern Medical Center at Dallas. He received a B.S. degree from California State University at Los Angeles, and a Ph.D. degree from the University of Wisconsin–Madison, where Muttaiya Sundaralingam was his advisor. His postdoctoral training was at the University of Alberta, Edmonton, with Robert Fletterick. Again with Dr. Fletterick, he was an assistant research biochemist at the University of California, San Francisco, before assuming his present position.

CELLS communicate with each other by secreting growth factors, cytokines, or hormones into the extracellular space, and by responding to factors produced by other cells. These chemical messengers act by changing patterns of gene expression within target cells, thereby altering the cells' metabolic or developmental program. Among the major unsolved questions in biology is how cells specifically recognize these chemical messengers and how the messages are transduced within cells.

We are using x-ray crystallographic techniques to define the molecular nature of the interaction between specific growth factors and their cognate receptors. Receptors are themselves protein molecules distributed on the outer surface of the plasma membrane surrounding the target cell. The goal of our research is to learn how receptor molecules specifically recognize growth factors, and how, if at all, the receptors alter their structure and chemical properties as a consequence of their interaction with the factors.

Tumor Necrosis Factors and Fibroblast Growth Factors

Tumor necrosis factors (TNF) are produced by cells of the immune system. Macrophages—the white blood cells responsible for engulfing cellular debris—produce large quantities of TNF-α (also known as cachectin) when stimulated by toxins carried on the surface of bacteria, as occurs in an infection. The avid attachment of this cytokine to receptors present on a variety of cells in the body triggers a series of events that mediate inflammation, endotoxic shock, and the wasting phenomenon, cachexia, from which the cytokine takes one of its names.

A related cytokine, TNF-β, is produced by T lymphocytes and has many properties in common with TNF-α, including the ability to bind to the same receptors. Both molecules are toxic to many types of tumor cells.

Two different types of TNF receptor have been found which, surprisingly, bear little amino acid sequence identity to each other. Both receptors, however, contain a structural "motif" composed of four inexact copies of a repeat rich in the sulfur-containing amino acid cysteine. Preliminary data from our own and other laboratories suggest that each trimer of TNF-α interacts with three receptor molecules. We hope to learn how these distantly related molecules interact with the same receptors and how this interaction might mediate a cellular response.

The three-dimensional atomic structures of TNF-α, and most recently of TNF-β, have been determined by Michael Eck, a graduate student in the laboratory (and now a Research Associate with Don Wiley and Stephen Harrison at HHMI, Harvard University). Despite the differences in amino acid composition and sequence, TNF-α and TNF-β have the same three-dimensional structure and assemble to form trimers composed of three identical protein subunits packed about a threefold axis of symmetry. However, because of the differences in the amino acid composition of the two cytokines, the surface of the TNF-α trimer is chemically quite different from that of the TNF-β trimer. We hope to determine how the same receptor can recognize these different molecular surfaces with equal affinity.

More importantly from a pharmacological perspective, we would like to be able to design TNF molecules that interact with one, but not the other receptor, or receptor antagonists that interact with only one of the two TNF species. Toward this goal, we are now attempting to determine the structure of the TNF-receptor complexes. The laboratory has been successful in growing crystals of the TNF-binding fragment of one of the two receptors, and crystallographic studies are now under way. We are also attempting to produce crystals of the complex formed between TNF-α and one of the receptors in order to study directly the interaction between the two molecules.

Last year we reported the three-dimensional structure of basic fibroblast growth factor (bFGF), one of a group of seven structurally related proteins that promote cell division. This factor is present in the space between cells, particularly in endothelial tissue, where it stimulates the movements of cells to sites of tissue genera-

tion or repair. Our work has lately focused on the interaction between bFGF and heparin, a complex, negatively charged sugar polymer that coats endothelial cells and is required to induce the mitogenic response that bFGF elicits. In a collaboration with Phillip Barr of Chiron, Inc., the laboratory is also attempting to determine the structure of the receptor for bFGF.

The signals generated by the engagement of extracellular messenger molecules with their receptors are, in many cases, transmitted across the plasma membrane to members of the G protein family. For example, when epinephrine binds to the β-adrenergic receptor, an intracellular protein called Gsα is induced to bind the nucleotide GTP and discard a pair of regulatory subunits. The GTP-bound Gsα is then an activator of a membrane-bound enzyme that catalyzes the synthesis of the intracellular messenger molecule cyclic AMP. In collaboration with Alfred Gilman (University of Texas Southwestern Medical Center at Dallas), we are undertaking crystallographic studies of Gsα and the related protein Giα to learn how these proteins may interact with other components of the signal transduction system. Research Associate David Coleman has obtained small crystals of the complex between Giα and a nonhydrolyzable GTP analogue, and we hope soon to resolve the structure of this fascinating molecule.

Glycogen Phosphorylase

A second focus of the laboratory has been to understand how a certain class of complex biological catalysts, called allosteric enzymes, regulate their own activity. An example of such an enzyme, studied in our laboratory, is glycogen phosphorylase. This molecule catalyzes the breakdown of a storage carbohydrate called glycogen into sugar units that can be used directly by the body to fuel muscle contraction or to maintain constant levels of glucose in the blood. The activity of this enzyme increases when it binds to "cellular messenger" molecules such as adenosine monophosphate, which signals an energy deficit in the cell, and decreases when it binds glucose, which signals an energy surplus. Glycogen phosphorylase can be chemically modified (phosphorylated) by other enzymes in response to hormonal signals (epinephrine), which also increases the catalytic activity of this enzyme. Our goal is to understand the molecular mechanics of the process by which catalytic machines such as phosphorylase can alter their activity in response to the cellular messengers.

We have recently determined the three-dimensional structure of active glycogen phosphorylase with the activator AMP bound to a regulatory site in the molecule. In comparing this structure with the inactive conformation of the molecule determined in Robert Fletterick's laboratory (University of California, San Francisco), we have learned how the three-dimensional structure of the molecule is altered in such a way as to increase the affinity of the enzyme for glycogen.

Insulin and the Islets of Langerhans

Donald F. Steiner, M.D.—Senior Investigator

Dr. Steiner is also A. N. Pritzker Distinguished Service Professor of Biochemistry and Molecular Biology and of Medicine at the University of Chicago Pritzker School of Medicine. He received his M.D. degree at the University of Chicago. His interest in insulin developed during postdoctoral training with Robert Williams at the University of Washington School of Medicine. After joining the faculty at the University of Chicago, he studied insulin action in the liver and, later, insulin biosynthesis. This work led to his discovery of proinsulin and preproinsulin. Dr. Steiner has received many honors, including the Lilly and Gairdner Awards, the Wolf Prize in Medicine, and several honorary degrees.

INSULIN is essential for normal growth and utilization of food. Diabetes, a disease due to insulin deficiency or defects in its action, is characterized by high blood sugar and such complications as blindness, heart disease, stroke, and increased susceptibility to infections. It affects 2–3 percent of people in developed countries.

Diabetes can be controlled through various combinations of diet, oral hypoglycemic agents, and/or insulin injections, depending on the type and severity of the disease. Such therapies, however, are often less than fully satisfactory because they may only retard the development of complications. A better understanding of how insulin is formed and secreted in a regulated manner into the bloodstream and how it acts on tissue receptors to control metabolism and growth are vitally important to the development of new therapeutic approaches.

Insulin is only made in the islets of Langerhans —small clusters of cells dispersed throughout the pancreas. Specialized islet cells also secrete other hormones that influence metabolism, including glucagon, somatostatin, amylin (or islet amyloid polypeptide), and pancreatic polypeptide. The islet hormones, like other regulatory peptides in the body, are derived from larger proteins called preprohormones.

These precursors contain additional portions that may guide them along special intracellular pathways, where they are concentrated into storage vesicles and processed into their biologically active forms. The contents of these vesicles are then released into the bloodstream in varying proportions to meet physiological requirements. One goal of our research is to learn more about how newly formed prohormones are separated from other proteins in the cell, concentrated into secretory granules, and processed into active hormones by highly specialized enzymes before being secreted.

Insulin secretion from β-cells in response to elevated plasma glucose is a complex electrochemical process resembling the transmission of nerve impulses or the contractions of the heart. In the insulin-producing β-cells, a specialized sensor mechanism couples the metabolism of glucose to ion channels in the plasma membrane. These channels, upon membrane depolarization, allow the selective entry of calcium into the cell, triggering the release of secretory granule contents. Certain oral hypoglycemic agents used to treat diabetes—the sulfonylurea drugs—appear to stimulate insulin secretion by inhibiting a specialized potassium channel in the β-cell membrane, which then initiates electrical depolarization of the cell. We are trying to learn more about the structure of this and other important ion channels in insulin-producing cells to understand both their normal functions and their possible malfunction in some forms of diabetes.

We also are studying mutations that affect insulin or proinsulin structure. Some of these occur in families and are associated with mild diabetes as a result of the synthesis of abnormal insulin molecules with greatly reduced biological activity. Other mutations in the insulin gene primarily affect the conversion of proinsulin into insulin, leading to elevated proinsulin in the circulation.

Insulin acts on tissues by binding to a large and complex protein receptor that is present on the surface of most cells, activating a tyrosine kinase that alters many intracellular processes through a cascade of intracellular phosphorylations. Insulin binding to the receptor also leads to the uptake and degradation of the hormone in the liver and other tissues. This process, known as receptor-mediated endocytosis, plays an important role by rapidly removing insulin from the circulation. By studying inherited defects in proinsulin and insulin receptor molecules, we are learning more about the normal processes of islet hormone production and action and how their derangement can lead to disease.

Prohormone-converting Enzymes

We have recently identified two cDNAs, which we call PC2 and PC3, in neuroendocrine cells. These encode proteases having catalytic domains similar to that of Kex2, a yeast prohormone-

processing enzyme. PC2 is more abundant in islet β-cells, while PC3 predominates in the anterior pituitary. Expression of PC2 in *Xenopus* oocytes gives rise to a 68-kDa protein that is active on substrates having characteristic sequences for prohormone cleavage recognition (i.e., Lys-Arg or Arg-Arg). The proteolytic activity requires calcium ions and is highest at pH 5.5. Other expression studies carried out in collaboration with Gary Thomas and his co-workers at the Vollum Institute indicate that PC2 and PC3 separately or together can appropriately process both proopiomelanocortin or proinsulin. Studies are now in progress on the genes encoding these proteases and their evolutionary origins, as well as on the synthesis, sorting, and activation of PC2 and/or PC3 in various neuroendocrine cells.

Regulation of Insulin Secretion

We have recently isolated and characterized cDNA and genomic clones encoding several voltage-dependent K^+ channel isoforms expressed in human islets or insulin-producing tumors. These channels are related to the *Shaker* family of *Drosophila* K^+ channels. Their electrophysiological characteristics have been examined by voltage clamp recordings of oocytes injected with synthetic mRNA. Efforts are also under way to identify and characterize an ATP-dependent potassium channel believed to play a key role in initiating β-cell depolarization in response to glucose or sulfonylureas.

Insulin Receptor Studies

Studies nearing completion have revealed an interesting and potentially important aspect of insulin proreceptor processing, namely that uncleaved proreceptors having exon 11 (which encodes 12 amino acids near the carboxyl terminus of the α-subunit) are fully functional, while those lacking this exon (due to alternative splicing of the mRNA) are markedly reduced in their affinity for insulin. Since most cells express receptors without exon 11, failure to cleave the proreceptor (as occurs in a family we have studied that has mutant proreceptors that cannot be processed) leads to severe insulin-resistant diabetes. The liver, kidney, islet β-cells, and placenta are the only tissues that express insulin receptors with exon 11, suggesting that alternative splicing may be physiologically relevant.

Biosynthesis of Islet Amyloid Polypeptide (IAPP)

IAPP, or amylin, a recently discovered product of the β-cell, is a peptide related to CGRP (calcitonin gene-related peptide) that is found in amyloid deposits in the islets of elderly diabetics and may play a role in impairing β-cell function. We have characterized the gene and cDNAs encoding the precursor of IAPP in humans and several other mammalian species. We are currently studying a transgenic mouse model to identify factors that may contribute to amyloid deposition (in collaboration with Niles Fox of the Lilly Research Laboratories).

Evolution of Insulin and Insulin-like Growth Factors

We have used the polymerase chain reaction to identify genes encoding insulin or the closely related insulin-like growth factors IGF-I and -II in primitive vertebrates. The identification of a hybrid insulin/IGF molecule in amphioxus, a protochordate, suggests that the insulin-like growth factors diverged from an ancestral preproinsulin-like protein in the very earliest stages of vertebrate evolution (about 600 million years ago). Studies nearing completion on the identification and structure of insulin and IGF receptors in these lower forms also lend support to this model for the origin of the insulin-like growth factors as unique vertebrate growth regulators.

Autoantibody Probes for Mammalian Gene Expression

Joan A. Steitz, Ph.D.—Investigator

Dr. Steitz is also Professor of Molecular Biophysics and Biochemistry at Yale University School of Medicine. She received her Ph.D. degree in biochemistry and molecular biology (with James Watson) from Harvard University and did postdoctoral work at the Medical Research Council Laboratory of Molecular Biology (with Frances Crick) in Cambridge, England. Her many honors include the Passano Foundation Young Scientist Award, the Eli Lilly Award in Biological Chemistry, the U.S. Steel Award in Molecular Biology, the National Medal of Science, the Dickson Prize for Science, the Warren Triennial Prize (shared with Thomas Cech), and the Christopher Columbus Discovery Award. Dr. Steitz is a member of the National Academy of Sciences.

KNOWLEDGE gained from basic research in the biomedical sciences sometimes provides answers to key questions in clinical medicine. Occasionally, however, the path is reversed and clinical studies provide information or materials that help to unravel basic biological processes. An example of this is our use of sera from human patients to determine the roles of previously mysterious small particles in normal cells. Particles called small nuclear ribonucleoproteins (snRNPs: pronounced "snurps") are found in the nucleus of the cells of humans and other higher organisms. Each snRNP is a tight cluster of one or more proteins with a small RNA molecule. SnRNPs are abundant in virtually all human cells and are remarkably similar among various mammalian species, suggesting that the particles must play important cellular roles.

Systemic lupus erythematosus (SLE) is one of a number of diseases in which the immune system mistakenly makes antibodies against the body's own molecules. Curiously, molecules that are very abundant in cells and highly conserved in evolution, such as DNA, are the most common targets of autoimmunity. Thus SLE patients often make autoantibodies against snRNPs.

Using SLE patients' antibodies to probe both the structures and functions of snRNPs, we have investigated the roles of various kinds of snRNPs in gene expression. These investigations began in 1979 when studies by Michael Lerner (then an M.D./Ph.D. student; now HHMI, Yale University School of Medicine) led to the hypothesis that the most abundant snRNP in mammalian cells, called the U1 snRNP, might be involved in RNA splicing, an early step in gene expression. In the making of a gene's product, its information coded in DNA is transcribed into an RNA copy called pre–messenger RNA, which is then "processed" into mRNA to direct the synthesis of a protein. The DNA and the pre-mRNA contain segments called exons, which code for the gene's product, and segments called introns, which are intermittent noncoding regions. Before leaving the cell nucleus as mRNA, the pre-mRNA is cut, the exons are spliced together, and the introns discarded. The individual exons must be precisely joined in the order they originally had in the gene. Sometimes differences in the way exons are spliced can lead to anomalous protein products in various tissues.

Evidence that snRNPs play central roles in pre-mRNA splicing has been obtained in several types of experiments, including use of autoantibodies from SLE patients to inhibit splicing in active cell extracts. We now know that participation of the most abundant snRNPs in mammalian cells (the U1, U2, U5, and U4/U6 particles) is essential and that splicing requires assistance from the snRNP proteins as well as their RNA molecules. SnRNPs recognize the splice junctions and the so-called intron branch point (where an unusual RNA structure is formed as an intermediate in splicing) and then assemble to align the exon ends so that precise splicing can occur. In this sense, snRNPs are much like the ribosomal subunits (also containing both RNA and protein) that assemble onto an mRNA to translate it into protein.

Current efforts in splicing are directed at understanding exactly how RNA-RNA interactions in the active splicing body (called the spliceosome) contribute to catalysis. Here, we are using two types of crosslinking approaches to identify contacts between the pre-mRNA molecule and either proteins or snRNAs in the assembled spliceosome. Novel crosslinks are now being analyzed that suggest how a cut-off exon may be held in the spliceosome for subsequent ligation. Also evolving is an increased understanding of how the spliceosome is related to some "self-splicing" introns, which are removed without proteins or other factors.

Mammalian cells also contain many minor snRNPs that are closely related to the splicing snRNPs. One is the U7, which is only 1/1,000 as abundant as the splicing snRNPs. We have recently demonstrated that it participates in forming the 3′ ends of histone mRNAs by using base-pairing to recognize a specific sequence in the

pre-mRNA just downstream of the cut site. Other related low-abundance snRNPs containing U11 and U12 RNA interact to form a two-snRNP complex, but their function is unknown.

A new autoantibody directed against the U11 snRNP has recently been found in the serum of a patient with scleroderma. Also in this group of related snRNPs, which probably all function in some aspect of mRNA maturation, are viral snRNPs. For instance, in marmoset cells infected by *Herpesvirus saimiri*, which causes malignant transformation of the T cells, the viral genome encodes five small RNAs. Current results suggest that these viral snRNPs may act to slow the degradation of mRNAs coding for oncoproteins or growth factors, thereby enhancing the process of cellular transformation.

Another type of patient autoantibody is directed against a different class of small RNPs localized in the nucleolus, where ribosomal RNA processing and assembly occur. The most abundant nucleolar snRNP (containing U3 RNA) is essential for the first step of ribosomal RNA processing. Recent analyses have revealed how U3 binds to the ribosomal pre-RNA, making specific base pairs. Separate studies are dissecting the signals that dictate the delivery of the U3 and related snRNPs to their nucleolar site of action.

Yet another autoantibody type precipitates EBERs, two small RNAs specified by Epstein-Barr virus (EBV), the causative agent of infectious mononucleosis, also implicated in several human cancers. Since EBERs are among the few viral products that are expressed in EBV-transformed cells, they must be important to the induction or maintenance of the transformed state. A highly abundant, highly conserved cell protein that binds the EBERs appears to reside in the cell's nucleolus. We hope that its further characterization will lead to an elucidation of EBER function.

Thus autoantibodies are potent probes for deciphering some of the fundamental reactions occurring in all mammalian cells, those involved in gene expression. Characterization of new cellular particles like snRNPs is significant for future studies of basic cellular processes and their alteration by disease. Furthermore, our research has provided new ways of diagnosing patient autoantibodies, which are helpful in the diagnosis and treatment of diseases like SLE.

Structural Studies of Protein–Nucleic Acid Interactions

Thomas A. Steitz, Ph.D.—Investigator

Dr. Steitz is also Professor of Molecular Biophysics and Biochemistry and of Chemistry at Yale University. He received a B.A. degree in chemistry from Lawrence College in Appleton, Wisconsin, and a Ph.D. degree in molecular biology and biochemistry from Harvard University, with W. N. Lipscomb. After a postdoctoral year at Harvard, he moved to the MRC Laboratory of Molecular Biology in Cambridge, England, as a Jane Coffin Childs fellow, with D. M. Blow. He next joined the Yale faculty, where he has remained, except for sabbatical work with K. Weber in Göttingen, West Germany, with A. Klug at Cambridge, and with J. Abelson at the California Institute of Technology. He has received the Pfizer Prize of the American Chemical Society and is a member of the National Academy of Sciences.

OUR general long-term goal is to determine the detailed molecular mechanisms by which those proteins and nucleic acids that are involved in the central dogma of molecular biology (DNA replication, transcription, and translation) achieve their biological functions. Virtually all aspects of the maintenance and expression of information stored in the genome involve interactions between proteins and nucleic acids. We are seeking to provide a structural and chemical basis for these fundamental processes.

Synthetase-tRNA Complex

Enzymes called aminoacyl-tRNA synthetases translate the genetic code by attaching the correct amino acid to a tRNA containing the appropriate anticodon. Of significant current interest is how these synthetases can accurately distinguish among the 60 or so similar tRNA molecules. Furthermore, how does RNA recognition by a protein differ from DNA recognition? Finally, can the structure of the present-day synthetase-tRNA complex provide any insights into the evolution of this central process and the evolution of the genetic code itself?

We have determined the crystal structure of glutaminyl-tRNA synthetase (GlnRS), a 64,000-molecular-weight monomeric protein, complexed with $tRNA^{Gln}$ and ATP. The GlnRS consists of four domains arranged to give an elongated molecule that interacts with the inside of the L-shaped tRNA from its anticodon to its acceptor end. GlnRS specifically recognizes the correct tRNA by interactions with the three bases of the anticodon and with base pairs of the amino acid acceptor stem of the tRNA. The three bases of the anticodon are unstacked and splayed out; each base binds into a separate recognition pocket on the enzyme. The extensive hydrogen-bonding interactions between the protein and the anticodon bases make the enzyme specific for the two glutamine anticodons (UUG and CUG) but none of the other 62 possible anticodons. The structural bases of Gln-tRNA synthetase recognition are currently being pursued by determining the structures of mutant tRNAs complexed with the enzyme.

The protein domain that contains the active site has a structure similar to that of the homologous domains of the tyrosyl- and methionyl-tRNA synthetases and by amino acid sequence similarity is homologous to another 10 of the 20 synthetases. Synthetases for 10 amino acids belong to a second unrelated class of synthetases. This work is supported by a grant from the National Institutes of Health.

Regulation of Gene Expression

In *Escherichia coli* a reduction in glucose concentration results in a rise in the levels of a second messenger molecule, cAMP, and subsequently to an increase in the proteins that metabolize other sugars. This is achieved because cAMP binds to the catabolite gene activator protein (CAP), which in turn binds to specific sequences at transcription start sites, activating the transcription of the catabolite genes. We wish to know how the binding of cAMP promotes the sequence-specific DNA binding of CAP and how this binding then activates the transcribing enzyme RNA polymerase.

We have now determined the structure of CAP cocrystallized with both a 30-bp DNA fragment and cAMP. The earlier CAP · cAMP structure had shown each subunit of this dimer to consist of two domains, the larger of which binds cAMP. The two small domains are seen to bind DNA, with the helix-turn-helix interacting in the major groove as anticipated. Strikingly, in the first complex the DNA is bent, with an overall bend of about 90°. In both complexes most of the bend is achieved by two large kinks of about 43° each. The relationships between this CAP-induced DNA bend and transcription activation are presently being pursued by attempts to crystallize CAP with polymerase and DNA.

Replication of DNA

E. coli DNA polymerase I functions primarily in the repair of DNA but is homologous to polymerases involved in replication. We have determined the structure of the Klenow fragment, a portion of Pol I that retains the polymerase and a 3′- to 5′-editing exonuclease activity. We have shown that a larger structural domain, which has a cleft sufficient in size to bind duplex DNA, contains the active site for the polymerase reaction, whereas a smaller domain has the active site for the exonuclease activity. Using site-directed mutagenesis, we have made an enzyme devoid of the editing exonuclease activity and determined its structure. We have grown two crystal forms of this protein complexed with a small DNA substrate. A high-resolution structure of one crystal form shows a single-stranded tetranucleotide bound to the exonuclease active site and 11 base pairs of duplex DNA bound to a cleft that runs at right angles to the major cleft. There are changes in the structure of the polymerase domain, including the movement of a thumb-like structure. To access the polymerase active site in the cleft, it appears that the DNA will have to bend by 90°. These structures begin to address the issues of how these two active sites work together on the same DNA substrate and how they both function to enhance the DNA-copying fidelity of this and other polymerases. This work is supported by a grant from the American Cancer Society.

Genetic Recombination

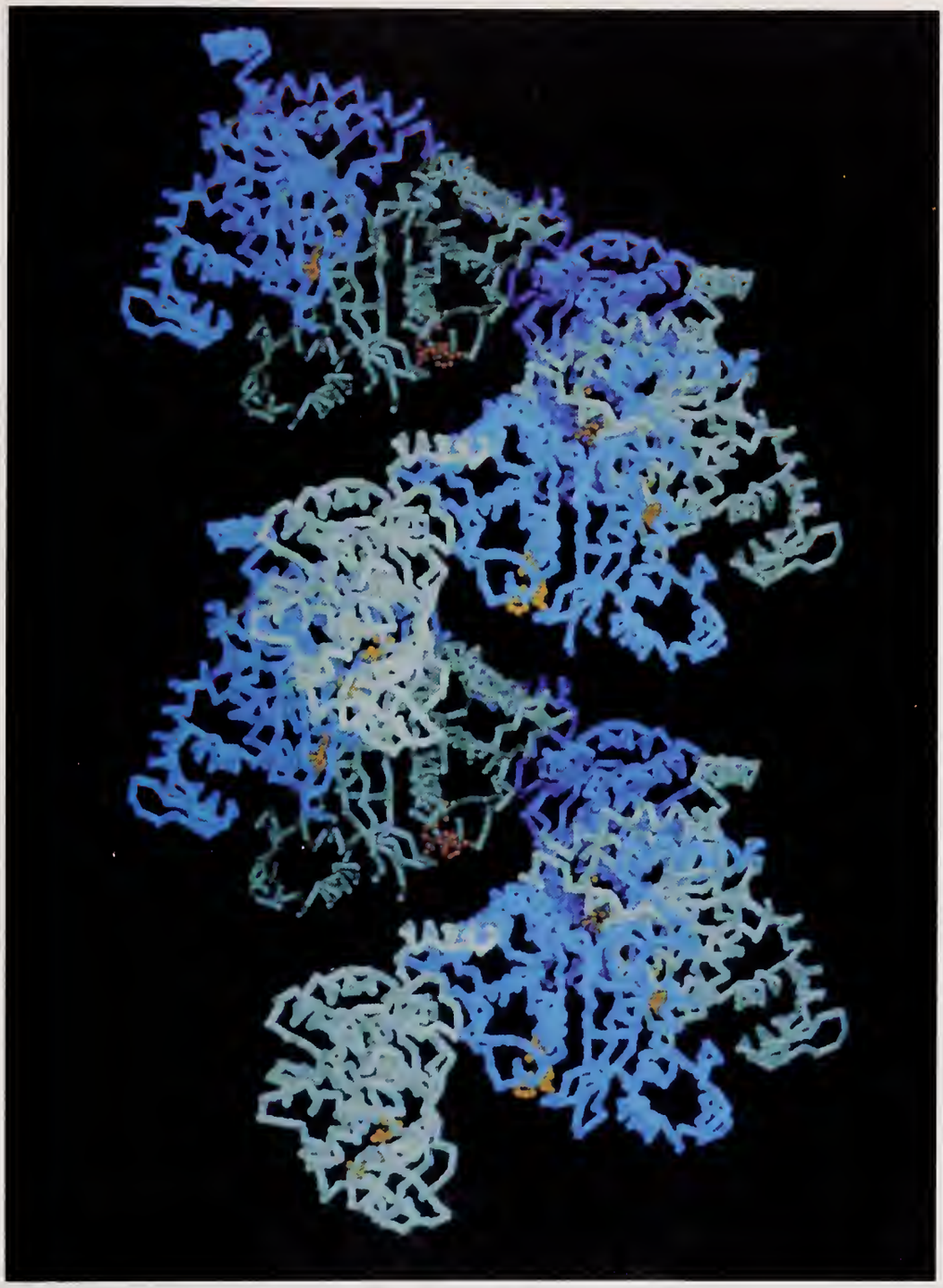
We have recently determined the crystal structures of two proteins that are involved in genetic recombination. One protein, called resolvase, catalyzes a site-specific recombination between two duplex DNAs of identical sequence. Resolvase is the product of a transposable element (a jumping gene) that can move throughout the *E. coli* population spreading drug resistance genes. This protein can bind to a specific duplex DNA sequence, align two DNA segments having the same sequence, cleave the two DNA duplexes, rearrange the duplexes, and re-ligate them, resulting in a recombinational event. We have determined the structure of the catalytic domain of this enzyme at 2.5-Å resolution. This structure helps to explain the phenotypes of many resolvase mutant proteins. This structure and that of the intact protein determined at 3-Å resolution provide ideas for a possible recombination mechanism. We have recently cocrystallized this protein with a 31-bp fragment containing the recombination site, whose structure should provide further clues to the mechanism of this reaction.

*E. coli rec*A protein is essential for general recombination in *E. coli*. Using the energy of ATP hydrolysis, *rec*A protein promotes the pairing of homologous duplex DNAs in preparation for recombination. The structure of *rec*A protein has been refined at 2.3-Å resolution. The subunit forms a helical filament in the crystal very similar to that formed on DNA and thus enables us to understand the many mutant *rec*A proteins made during the past decade and relate its structure to its functions in nucleotide binding, DNA binding, and the SOS response. Our goal is to understand how ATP hydrolysis and the homologous pairing of DNA are coupled. We have now produced good crystals of the next enzyme in the pathway of recombination, ruvC, which cleaves the DNA recombination intermediate called the Holliday structure. The work on resolvase and the *rec*A protein is supported by a grant from the National Institutes of Health.

HIV Proteins

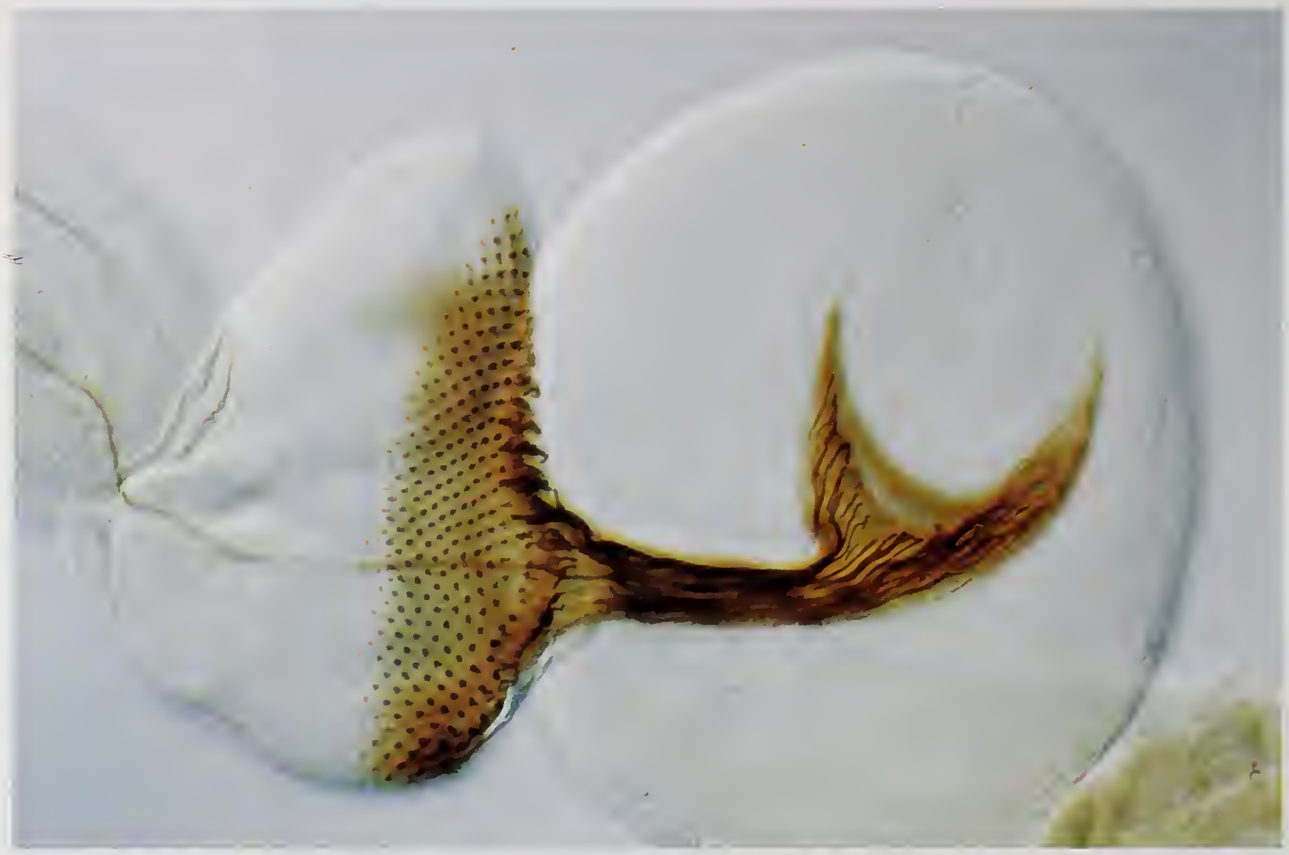
During the past year we have determined the structure of human immunodeficiency virus (HIV) reverse transcriptase complexed with a nonnucleotide inhibitor that shows promise as an anti-AIDS drug. The enzyme is a heterodimer of a 66-kDa polypeptide containing a polymerase and RNase H domain and a 51-kDa polypeptide containing only the polymerase domain. This dimeric polymerase shows an asymmetric structure with one active-site cleft running from the p66 domain to the RNase H domain. The p51 subunit has no cleft. The similarities between the catalytic domains of HIV reverse transcriptase and Klenow fragment suggest that all polymerases have the same catalytic site. We have been successful in diffusing other nonnucleotide inhibitors into the crystal and find that they bind into a deep pocket near to, but not overlapping with, the polymerase catalytic site. We expect that design of additional inhibitors based on the structure will be possible.

Recently a peptide fragment of the transactivator protein Tat has been cocrystallized with a fragment of the RNA site to which it binds, TAR. This work is supported in part by a grant from the National Institutes of Health.



A representation of the crystal structure of Escherichia coli rec*A protein. This protein catalyzes DNA strand exchange in general genetic recombination. Shown are 12 subunits of the crystal in* α*-carbon backbone representation, with adjacent subunits alternately light and dark blue. A bound ADP molecule is represented in orange.*

Figure by Randall Story and Thomas Steitz from research reported in Story, R.M., Weber, I.T., and Steitz, T.A. 1992. Nature *355:318–325.*



Visualization of photoreceptor axons in the developing visual system of Drosophila. *The axon bundles emerge from each ommatidium of the developing retina and project through an epithelial tube, the optic stalk, to specific retinotopic positions in the brain. Guidance of these axons relies largely on positional information.*

Research of Hermann Steller.

Pattern Formation and Neuronal Cell Recognition in the Drosophila *Visual System*

Hermann Steller, Ph.D.—Assistant Investigator

Dr. Steller is also Associate Professor of Neurobiology at the Massachusetts Institute of Technology and Adjunct Assistant Neurobiologist at Massachusetts General Hospital, Boston. He was born in the Federal Republic of Germany and received a Diplom in biology from the Johann-Wolfgang-Goethe University, Frankfurt. His graduate studies were done with Vincenzo Pirrotta at the European Molecular Biology Laboratory and with Ekkehard Bautz at Heidelberg University. His postdoctoral work was done with Gerald Rubin in the Department of Biochemistry at the University of California, Berkeley. Dr. Steller is currently also a Searle Scholar and a Pew Fellow in the Biomedical Sciences.

THE overall objective of our research is to understand how functional neuronal circuits are established and maintained during development. Our current work is focused on three major areas.

Axon Guidance and Neuronal Cell Recognition

We are studying two different stages of visual system development to investigate the cellular and molecular mechanisms by which axons find and recognize their proper synaptic partners. The optic nerve of the *Drosophila* larva is a simple, well-described model system. Connectivity defects of the larval optic nerve can be rapidly and reliably detected in mutant embryos by staining with neuron-specific antibodies. In addition, a simple behavioral test, larval phototaxis, provides an assay for functional connections of the larval optic nerve. This permits systematic screening for mutants with abnormal axonal projections, which can be subsequently analyzed in detail with respect to defects in axonal guidance, target recognition, and synapse formation.

We have previously identified a gene, *disconnected* (*disco*), which is required for establishing stable connections between the larval optic nerve and its target cells in the developing brain. We have cloned the *disco* gene and determined its structure, nucleotide sequence, and pattern of expression. These studies suggest that *disco* encodes a transcription factor with autoregulatory properties. Consistent with such a function we have recently found that *disco* protein has sequence-specific DNA-binding activity *in vitro* and that two high-affinity binding sites are located very close to the *disco* transcription unit. Ectopic expression of *disco* protein under an inducible promoter in transgenic flies results in severe developmental defects and embryonic lethality. These defects include a drastic reduction of the axon scaffold and connectivity defects in both the peripheral and central nervous systems. We are now testing the idea that *disco* regulates the expression of cell adhesion and/or cell recognition molecules that are required for the establishment of stable connections between the larval optic nerve and its target cells in the brain.

More recently we have begun to study axon guidance and neuronal cell recognition in the adult visual system. The compound eye of *Drosophila* consists of approximately 800 repeating units called ommatidia. Each ommatidium contains eight photoreceptor neurons, which represent three major cell types that project to different target cells in the optic ganglia. The major class of photoreceptors, R1–6, establishes synaptic connections with neurons in the first optic ganglion, the lamina. Photoreceptor axons from R7 and R8 project deeper into the brain to different regions of the second ganglion, the medulla. Early during visual system development, all eight photoreceptor axons from each ommatidium grow as a bundle to specific retinotopic positions in the developing brain. The growth cones of these axons have to navigate over a long distance and make a number of highly specific choices.

We would like to understand what signals guide axons to their proper destinations and how these signals are generated, received, and interpreted. To address these questions, we have screened for mutations that perturb the projection pattern of photoreceptor cells at very early developmental stages, when axons enter the brain. We have found a number of mutants with severely abnormal patterns of axon ingrowth. The developmental and genetic characterization of this material is in progress.

Role of Innervation for Neurogenesis and Survival of Target Cells

It has been noticed for many years that synaptic input can have a profound influence on the fate and differentiation of target cells. Cell death in the absence of incoming projections is a dramatic example of how innervation can affect developmental decisions, and many neurological disorders are thought to arise from defective interactions between neurons and their targets. In

Drosophila the proper development of the adult optic ganglia, the central nervous system portion of the visual system, depends on innervation from the eye. In the absence of retinal innervation, adult flies entirely lack the first optic ganglion, the lamina, which receives direct synaptic input from the outer photoreceptor cells R1–6.

We have found that the birth of lamina neurons is controlled by innervation from the developing eye. The arrival of photoreceptor axons in the brain induces a wave of mitotic activity that produces the lamina neurons. These results suggest a novel mechanism for matching the number of target neurons in the first optic ganglion to the number of incoming photoreceptor axons, and they explain how developmental synchrony between the *Drosophila* retina and first optic ganglion is achieved. We are now using several different approaches to elucidate the detailed cellular and molecular mechanisms underlying this process.

While the importance of retinal innervation on the development of the adult optic ganglia of *Drosophila* is well documented, little is known about retrograde effects of the brain on photoreceptor cells in the compound eye. We have recently discovered the first evidence for the existence of such retrograde effects in the *Drosophila* visual system. Although photoreceptor cells develop normally in the absence of connections to the optic ganglia, we have found that their continued survival requires these connections. This situation is reminiscent of trophic interactions that are commonly found in vertebrates.

Genetic Control of Cell Death

Apoptosis, the deliberate and orderly removal of cells by natural death, is a prominent feature of normal development throughout the animal kingdom. In many organisms, a large number of cells die in the absence of obvious external insults. For example, in vertebrates neurogenesis produces about twice as many neurons as are needed in the mature nervous system, and approximately half of these neurons are eliminated by cell death. We are interested in isolating genes that are required for the initiation or execution of cell death in *Drosophila*. We have found that the ultrastructural characteristics of cell deaths seen in the *Drosophila* embryo are strikingly similar to apoptotic deaths described in other systems. We have developed techniques utilizing the vital dyes acridine orange and nile blue that permit the rapid and reliable visualization of apoptotic cells in live embryos, and we have used these methods to screen for cell death–defective mutants. We have identified a complex genetic locus on the third chromosome that is required for either the commitment to or the execution of a cell death program. We have cloned the DNA encompassing this locus, and we expect that its molecular characterization will provide insight into the biochemical mechanisms underlying cell death in *Drosophila*, and possibly other organisms as well.

Molecular Genetics of Nematode Development and Behavior

Paul W. Sternberg, Ph.D.—Associate Investigator

Dr. Sternberg is also Associate Professor of Biology at the California Institute of Technology and Adjunct Assistant Professor of Anatomy and Cell Biology at the University of Southern California School of Medicine, Los Angeles. He received a B.A. degree in biology and mathematics from Hampshire College and a Ph.D. degree in biology from the Massachusetts Institute of Technology for work with Robert Horvitz. He did postdoctoral research in yeast molecular genetics with Ira Herskowitz at the University of California, San Francisco. Dr. Sternberg is also a Presidential Young Investigator.

USING the nematode *Caenorhabditis elegans*, our laboratory takes a molecular genetics approach to basic questions in developmental biology and neurogenetics: What are the molecular mechanisms by which cells interact to establish a spatial pattern of cell types? What is the genetic and cellular basis for morphogenesis? What establishes the asymmetry of individual cells? How are the instructions for innate behavior encoded in the genome? Our major strategy is to identify mutations that make cells or animals misbehave and then to study the functions of the genes defined by these mutations, using a combination of molecular cloning and genetic analysis. A second strategy is to clone nematode homologues of genes identified in mammals and then to elucidate the functions of those genes in nematodes.

In this past year we focused on the development and function of the *C. elegans* male spicules—innervated structures crucial to successful mating. Each of the two spicules comprises nine cells: two sensory neurons, one motoneuron, and six supporting cells. By studying spicule development, we have identified a new example of induction during nematode development. In the developing male, either of two cells signals spicule precursor cells to generate particular sets of spicule cells. This inductive signaling process requires the *lin-3* growth factor, the *let-23* tyrosine kinase, and the *let-60 ras* genes that we cloned over the last two years. *lin-3* encodes an inductive signal for the hermaphrodite vulva, and it is likely that it acts as an inductive signal for proper spicule development as well. Thus we have found that a cascade of proto-oncogenes specifies cell fates in several aspects of nematode development. In addition to this inductive signal, at least three other signals are also necessary for the correct specification of spicule precursor cells.

Because *C. elegans* hermaphrodites are internally self-fertilizing—each animal producing both sperm and ova—male mating and thus spicule function is dispensable. Thus mutant strains defective in male mating can be easily propagated and the mating process studied. We have used a simple behavioral test—the ability of males to sire progeny—to isolate mutants that are unable to mate. Some mutant males have obvious defects in the development of male-specific structures. Others, called Cod (for copulation defective), are anatomically normal yet defective in mating behavior.

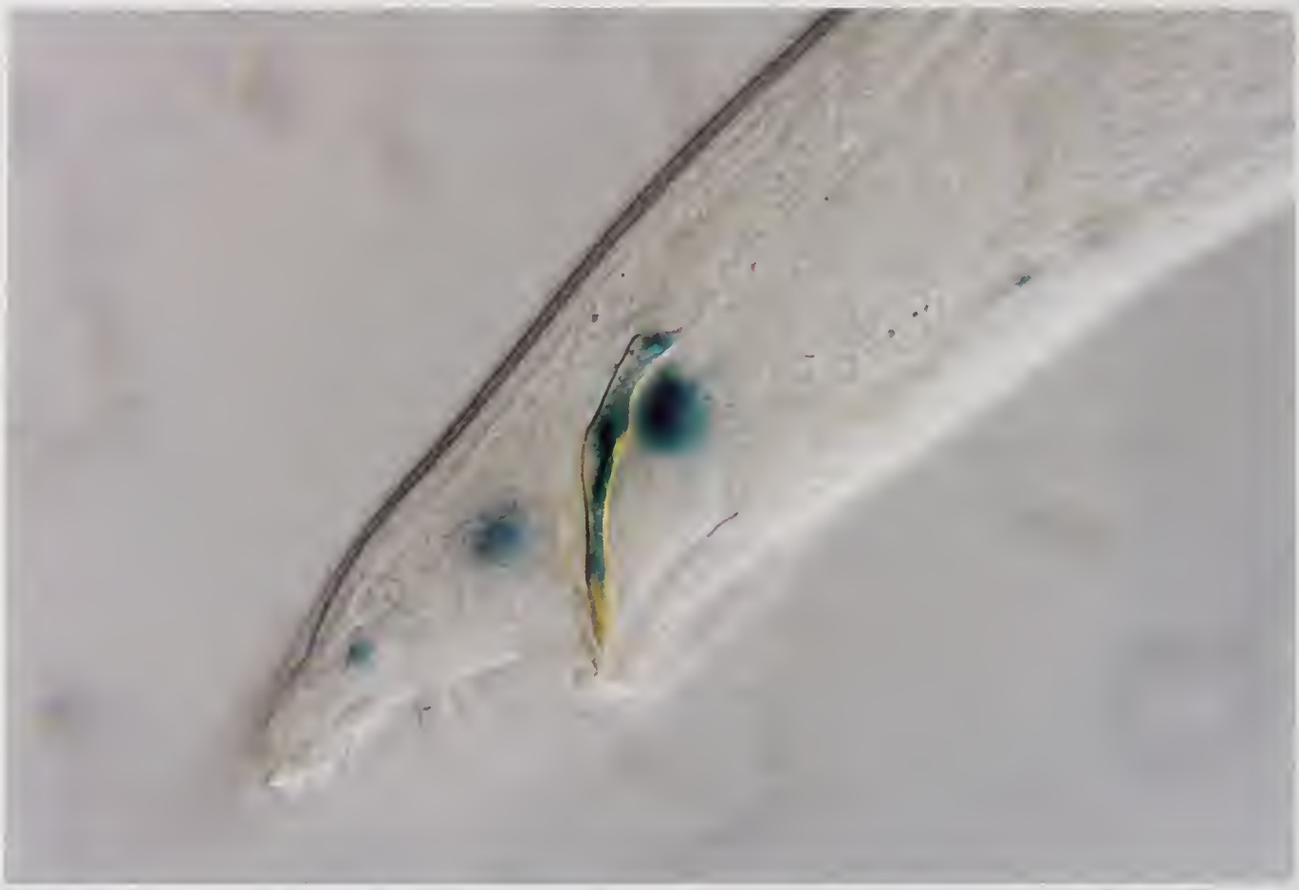
By studying the Cod mutants, we hope to elucidate how genes control each step in male mating behavior. We have isolated a set of mutants, have characterized the mating defect of each strain, and have begun placing these mutations on the genetic map. Most of the mutants analyzed are defective at only a single step in the mating process. These steps include 1) attraction to hermaphrodites, 2) maintaining contact with hermaphrodites, 3) location of the vulva, 4) insertion of spicules, and 5) transfer of sperm. For example, a mutant male defective in step 4 will locate the hermaphrodite vulva but fail to insert his spicule. Having mutants blocked at defined steps will allow us to identify genes necessary to specify this innate behavior.

To identify the cells responsible for each step in mating behavior, we kill individual cells with a laser microbeam and observe the consequences. For example, the spicule motoneuron and both spicule sensory neurons are necessary for spicule insertion. One of our G protein α-subunit genes is expressed in one of these neurons, suggesting that it might serve to regulate spicule insertion. Another G protein, homologous to human G_o, is expressed in male diagonal muscles, required for initial steps in mating. We have begun to test whether any of the Cod mutants are defective in these G protein genes.

The establishment of cellular asymmetry is a fundamental aspect of cell regulation. We have begun to study this problem in the context of the 2° vulval precursor cell. The *lin-18* gene (cell lineage gene number 18) is necessary for the asymmetry of the 2° vulval precursor cell. We have mapped the *lin-18* gene to a manageable region of the X chromosome. We have found that the 2° cells will orient posteriorly in the animal, unless they receive a signal from the developing gonad to orient anteriorly. This signal is distinct

from the inductive signal that instructs the 2° cells to be 2° and not other types of cells. The *lin-18* gene must be involved in the response to this signal.

In another project, we have been analyzing the function of the nematode homologues of signal transduction molecules identified in mammals. We have cloned a nematode gene similar to the *raf-1* protein kinase as well as four nematode genes similar to mammalian G protein α-subunit genes. We have placed these on the *C. elegans* genomic map. To identify the cells expressing these genes, we have constructed chimeric genes consisting of their transcriptional control regions linked to a reporter gene. We then introduced the engineered genes back into the nematode by microinjection of DNA to create transgenic nematodes. All four G protein genes are expressed primarily in neurons, including neurons in the male tail. We plan to engineer poisonous forms of these genes to disrupt their function in intact animals.



A G protein α-subunit gene expressed in a sensory neuron of a male spicule of Caenorhabditis elegans. *The cell body is seen as a large round blue spot, and the dendrite as blue inside the yellow spicule.*

Research of Paul Sternberg; photograph by Jane Mendel.

Why Do We Drink Coffee and Tea?

Charles F. Stevens, M.D., Ph.D.—Investigator

Dr. Stevens is also Professor of Molecular Neurobiology at the Salk Institute for Biological Studies and Adjunct Professor of Pharmacology and of Neuroscience at the University of California School of Medicine, San Diego. He received his B.A. degree in psychology at Harvard University, his M.D. degree at Yale University, and his Ph.D. degree in biophysics at the Rockefeller University for studies with Keffer Hartline. He was a member of the faculties at the University of Washington Medical School and at Yale Medical School before joining the Salk Institute. Dr. Stevens is a member of the National Academy of Sciences.

COFFEE and tea are widely used in many societies as mild stimulants. How do they work? Recent research in our laboratory provides an answer.

One of the more remarkable features of the brain is its ability to modify its own neuronal characteristics, a phenomenon known as *neuromodulation*. Neurons can cause other neurons to alter their properties, and thus can change the computations carried out by the modified neuronal circuits.

Neuromodulation occurs in many ways. One of the most important is the regulation of synaptic strength—the force and effectiveness with which one neuron transmits signals to another. Some changes in synaptic strength are very long lasting and are thought to underlie the storage of memories. Other modifications of synaptic strength occur more rapidly and constitute a moment-to-moment tuning up of circuit function. A major goal of our laboratory has been to elucidate the mechanisms through which synaptic strength is neuromodulated. In the course of this work, we have gained insights into the actions of caffeine and theophylline, the stimulants in coffee and tea, respectively.

The chemical adenosine is released at synapses together with neurotransmitters and is present in the fluids bathing neurons. Neurons display on their surface several different types of receptors for adenosine, which couple adenosine binding to second messenger cascades (such as the one involving cAMP). Adenosine is a very potent regulator of synaptic strength: concentrations above the usual levels result in large *decreases* in strength. Here is the interesting part: methylxanthines, a family of chemicals that includes caffeine and theophylline, bind to adenosine receptors and block their uptake of adenosine.

Some "fake" agonists that bind to a receptor will activate it as if they were natural agonists, but the methylxanthines do not do this. They occupy the receptor's adenosine binding site and thereby prevent adenosine from doing so. Rather than mimic the action of adenosine, the methylxanthines act as if taking it away. Thus they increase synaptic strength, indicating that the normal brain levels of adenosine are sufficient to produce and maintain a partial decrease in synaptic strength.

We examined the effect of adenosine itself, and of adenosine receptor blockers, on synaptic transmission in brain slices from the hippocampal region (specifically from dentate) of rats. Using the whole-cell recording method, a technique that permits high-resolution detection of currents that flow as a result of synaptic activation, we were able to determine the specific consequences of adenosine and methylxanthine action.

To interpret the effects of these drugs, one needs to know certain features of normal synaptic transmission. A synapse contains a number of microscopic membrane-bounded spheres, known as synaptic vesicles, whose contents are released from the cell by a membrane fusion event (exocytosis). Each of these vesicles contains a unit amount of neurotransmitter, such as glutamate, that binds to special receptors in the target neuron surface membrane and produces a signal. The strength of a synapse therefore depends jointly on the number of vesicles whose contents are released and on the size of the response produced by one vesicleful of the neurotransmitter. Because the size of the synaptic response must be integral—a multiple of that produced by a single vesicle—this single-vesicle response is called a quantum or quantal response. The quantal size depends, in turn, on the number of transmitter molecules contained in a vesicle and on the number and responsiveness of the receptors displayed by the target neuron.

Only a certain number of vesicles are available, of course, to release their contained neurotransmitter. This number is usually denoted by N. What is interesting and significant about the neurotransmitter release process is that it is probabilistic. That is, only a fraction of the N vesicles available to release their contents do so, and that number is determined the same way one would calculate the number of heads in N coin flips, with a probability p for heads on any flip. Thus

the strength of a synapse is determined by N, p, and the quantal size. There are techniques for measuring the quantal size directly, and a kind of statistical computation, known as quantal analysis, can estimate N and p.

We have carried out this procedure in the dentate neurons with either added adenosine or adenosine receptor antagonists. The quantal responded size was unaffected by any of the drugs. Quantal analysis, however, reveals a pronounced effect on the release probability: adenosine decreases and methylxanthines increase the chances that a vesicle will undergo exocytosis and release its neurotransmitter when a nerve impulse arrives at the synapse.

Thus methylxanthines antagonize baseline adenosine effects and strengthen synapses. Why does this act as a stimulant? About two-thirds of the brain's synapses are excitatory and the other third are inhibitory. Many drugs shift the balance between excitation and inhibition and thereby have a net stimulatory or depressive effect on the brain's activity. Barbiturates, alcohol, and benzodiazipines (e.g., Valium) all selectively increase the strength of inhibitory synapses and shift the balance from excitation toward inhibition. We find that the effects of adenosine and adenosine antagonists on synaptic transmission are limited to excitatory synapses. Apparently the inhibitory synapses lack adenosine receptors or the second messenger cascades that are necessary for the expression of their action.

In summary, coffee and tea act (as many have discovered independently) as antialcohol and antibarbiturate agents, in that they shift the brain's excitation/inhibition balance in the excitatory direction. They do this by antagonizing the effects of background levels of adenosine on synaptic strength. Their precise mechanism of action is to increase the probability of release of neurotransmitter molecules from synaptic vesicles.

Morphogen Gradients and the Control of Body Pattern in Drosophila

Gary Struhl, Ph.D.—Associate Investigator

Dr. Struhl is also Associate Professor of Genetics and Development at Columbia University College of Physicians and Surgeons. He received B.S. and M.S. degrees in biology from the Massachusetts Institute of Technology and his Ph.D. degree in genetics from the University of Cambridge, England. His graduate studies were carried out with Peter Lawrence at Cambridge. Before joining the faculty at Columbia, he conducted molecular and genetic studies on Drosophila *in the laboratory of Tom Maniatis at Harvard University. In addition to several fellowships, he received the McKnight Neuroscience Development Award.*

SINCE the birth of embryology as an experimental discipline, it has been apparent that the development of cell and body patterns depends on robust and complex systems of spatial information. Yet, until recently, we had little idea of the physical nature of such systems or the mechanisms by which they are generated or interpreted. Considerable attention has been given to the possibility that gradients of "form-generating" substances, or morphogens, might provide this information. For example, a localized diffusible morphogen could generate a gradient that would provide a series of concentration thresholds, each determining distinct cellular behaviors (e.g., the development of particular pattern elements) as a function of distance from the source.

Current research in this laboratory is directed toward identifying potential morphogen gradients and determining both how they arise and how they control pattern. Over the past 20–30 years, a variety of embryologic and genetic experiments have suggested that such gradients play a key role in controlling cell and body patterns in insects. Indeed, Christiane Nüsslein-Volhard and her colleagues have identified several such morphogens involved in specifying the basic body plan of the fruit fly *Drosophila*.

One of these determinants, the protein product of the gene *bicoid* (*bcd*), is clearly expressed as a gradient that peaks at the anterior pole of the early embryo and controls anterior body pattern (head and thorax). Another, *nanos* (*nos*), is required for generating posterior body pattern (abdomen), though its properties are less well understood. Our immediate goal during the past year or so has been to determine how these two systems work. As we describe below, the controlled expression of a single regulatory protein, *hunchback* (*hb*), has proved to be the key to understanding both molecular mechanisms.

Anterior Pattern: Transcriptional Activation of *hb* by *bcd*

Early in embryogenesis, the *hb* gene is activated in a broad domain extending from the anterior pole to a boundary halfway down the body. This event depends critically on the *bcd* morphogen. Moreover, it is essential for the generation of a large portion of anterior pattern, including all three thoracic segments.

In an initial series of experiments, we found that *bcd* behaves as a transcriptional regulator, capable of directly binding the *hb* gene and activating transcription in a remarkably sensitive, concentration-dependent fashion. Hence, when the concentration of *bcd* protein exceeds a critical threshold, it binds a series of adjacent sites in the DNA immediately upstream of the *hb* promoter and activates transcription; however, when the concentration of *bcd* protein falls beneath this threshold, binding does not occur and the *hb* gene remains silent. The concentration gradient of *bcd* protein therefore determines where the *hb* gene is transcribed by positioning this threshold along the anteroposterior axis.

In the process of examining how the *bcd* gradient controls *hb* gene expression, we created a series of hybrid genes that contain only some of the binding sites for *bcd* protein normally present upstream of the *hb* promoter. These hybrid genes, like *hb* itself, are activated in anterior portions of the body under the control of the *bcd* protein gradient. However, their domains of expression are abbreviated relative to that of *hb*, extending only a quarter or a third of the way down the body. These and other experiments suggest that the *bcd* gradient can trigger several spatially distinct responses, the posterior boundary of each depending on the affinity of *bcd* protein for a given target gene.

Recently several new genes have been identified that appear to be expressed in a series of overlapping anterior domains. These genes are required for generating particular portions of head and thoracic pattern and may constitute other direct targets of the *bcd* gradient.

Posterior Pattern: Translation Repression of *hb* by *nos*

Transcriptional activation of the *hb* gene by *bcd* is only one of two sources of *hb* mRNA in the

early embryo. The second source is maternal *hb* mRNA synthesized during oogenesis and deposited uniformly throughout the egg. These transcripts are translated only after fertilization, and their translation is repressed in more-posterior portions of the egg, owing to *nos* activity emanating from the posterior pole.

We recently found that the ability of *nos* to repress *hb* translation depends on the presence of two copies of a short cis-acting regulatory sequence in the 3′ noncoding portion of the *hb* mRNA. As in the case of the DNA sites mediating the binding and regulation of *hb* by *bcd* protein, the number and quality of the target sites in *hb* mRNA can determine the pattern of expression of *hb* protein.

However, in contrast to *bcd*, which has many other signaling roles, *nos* appears to have no role other than to block posterior expression of *hb* protein. Indeed, we have been able to show that *nos* activity is completely dispensable if the translation of maternally derived *hb* mRNAs is blocked by other means (such as mutation). Thus the control of posterior body pattern may depend solely on the repression of maternal *hb* transcripts by *nos*.

Posterior Body Pattern Is Controlled by *hb* Protein Acting as a Classical Gradient Morphogen

The *bcd* and *nos* determinants have opposite effects on *hb* protein expression. The *bcd* gene triggers the synthesis of *hb* transcripts, and hence protein, anteriorly; *nos* blocks the translation of both maternal and zygotic *hb* transcripts posteriorly. Together the actions of *bcd* and *nos* generate a graded pattern of *hb* protein in which the concentration declines from uniformly high levels in the anterior half of the body to undetectable levels in the posterior half.

As described above, our analysis of the interaction between *nos* and *hb* indicates that *nos* itself can be rendered dispensable, provided that maternal *hb* transcripts can be inactivated by other means. In these unusual embryos, *hb* protein is still expressed differentially along the anteroposterior axis, owing to zygotic activation of the *hb* gene by *bcd*. This finding, taken together with the critical role *nos* normally plays in repressing the translation of *hb* mRNAs, suggests that the distribution of *hb* protein may be the critical determinant of posterior body pattern.

We tested this possibility by generating embryos in which the differential expression of *hb* protein along the anteroposterior axis has been systematically altered while all other known signaling systems are eliminated or held constant. By examining the expression of the subordinate regulatory genes *Krüppel*, *knirps*, and *giant*, each of which is normally responsible for controlling a particular subdomain of thoracic and abdominal pattern, we have been able to show that the gradient of *hb* protein provides a series of distinct concentration thresholds that govern where these genes are expressed.

Thus *hb* appears to control posterior body pattern by acting as a classical gradient morphogen. In this regard, it functions as the posterior counterpart to the anterior morphogen *bcd*.

Secretory Pathways in Neurons

Thomas C. Südhof, M.D.—Investigator

Dr. Südhof is also Professor of Molecular Genetics at the University of Texas Southwestern Medical Center at Dallas. He received his M.D. degree and his doctorate from the Georgia Augusta University of Göttingen, FRG. He obtained postdoctoral training first with Victor Whittaker at the Max Planck Institute for Biophysical Chemistry, Göttingen, and then with Michael Brown and Joseph Goldstein in Dallas.

NEURONS communicate with one another by means of chemical signals. The known communication pathways between neurons are of several kinds:

- Fast, point-to-point transmission of signals between neurons occurs at the synapse.
- Long-lasting modulatory signals that often reach many cells are transmitted outside of synapses by neuropeptides and other mediators.
- Short-range diffuse signals are probably spread by lipophilic messengers, such as nitric oxide or arachidonic acid.

Of these pathways, signaling between neurons at the synapse is quantitatively the major form of cell-to-cell communication in the central nervous system. Synapses are abundant in the nervous system, and their activity provides the basis of brain function. However, the slow communication pathways are clearly an essential counterpart to the fast point-to-point signals. The coexistence of different signaling pathways in the same neuron increases the complexity of the neuronal networks. Brain function will clearly not be understood until we gain insight into the molecular mechanisms that govern these signaling pathways.

Work in our laboratory addresses the question of how nerve cells send out chemical signals. We are concentrating on the synapse as the most abundant signaling pathway. Here the chemical signals, the neurotransmitters, are prepackaged in unique cellular organelles called synaptic vesicles and released from the presynaptic neuron by secretion. This secretion is achieved by exocytosis, the fusion of synaptic vesicles with the synaptic cell membrane. After fusion the empty synaptic vesicles are quickly re-endocytosed and refilled with neurotransmitter. They become competent for secretion again in a short time, allowing the neurons to fire rapidly.

We have taken two avenues to the exploration of the molecular basis of signal transmission at the synapse. The first approach has been to study synaptic vesicles and their components as the central organelle in neurotransmitter release. The second approach consists of a characterization of the presynaptic plasma membrane as the point of signal release.

Characterizing the molecular components of synaptic vesicles constitutes a long-term project that we are largely carrying out in collaboration with Reinhard Jahn (HHMI, Yale University). This project has led to the molecular characterization of more than 10 synaptic vesicle proteins, which together account for approximately one-third to one-half of the total vesicle protein by mass. The goal of this project is twofold.

First, we would like to achieve a complete description of the synaptic vesicle in molecular terms. This is not only a precondition to any eventual understanding of how the synaptic vesicle pathway works, but would also provide the first molecular anatomy of an organelle.

Second, we would like to explore the functions of each vesicle protein in neurotransmitter release, using biochemical and genetic techniques. This part of the project has progressed to the point that interesting biochemical properties of several vesicle proteins have been elucidated. Moreover, the feasibility of mouse genetics to probe the functions of these proteins has been demonstrated. For example, we have recently shown that a synaptic vesicle protein named synaptotagmin binds Ca^{2+} and phospholipids in a ternary complex at physiologic Ca^{2+} concentrations. This result suggests a function for synaptotagmin in synaptic vesicle fusion, a possibility now being explored in transgenic mice.

Another approach we are pursuing to elucidate synapse function is the study of the presynaptic plasma membrane. The membrane serves two basic functions. It binds synaptic vesicles and fuses with them in a Ca^{2+}-dependent manner, thereby releasing neurotransmitter; and it contacts the postsynaptic site and aligns pre- and postsynaptic membranes with each other. Both functions are probably performed by specific protein components.

To identify a component of the presynaptic plasma membrane that may be involved in its functions, we have studied the receptor for a neurotoxin called α-latrotoxin. This protein, derived from venom of the black widow spider, binds spe-

cifically to the presynaptic membrane and causes massive neurotransmitter release. Purification and amino acid sequencing of the receptor for α-latrotoxin led to the discovery of a new family of neuron-specific cell surface proteins that we named neurexins. Their structure and localization suggest that they may be synaptic recognition molecules.

Surprisingly, we also found that the α-latrotoxin receptor interacts with a synaptic vesicle protein, synaptotagmin, whose Ca^{2+}-binding properties have already implicated it in exocytosis. These results raise the possibility that the neurexins may have a role in aligning the synapse both extra- and intracellularly. Further studies of the neurexins should allow us to gain new insights into how synapses are formed and neurotransmitters released. Results should be relevant for understanding brain function under normal and pathologic conditions.

Transcription Factors in Cell Growth and Kidney Differentiation

Vikas P. Sukhatme, M.D., Ph.D.—Assistant Investigator

Dr. Sukhatme is also Associate Professor of Medicine and of Molecular Genetics and Cell Biology at the University of Chicago. He received his Ph.D. degree in theoretical physics from the Massachusetts Institute of Technology. He then received his M.D. degree from Harvard Medical School. After residency and clinical fellowship training at the Massachusetts General Hospital, Boston, he completed his postdoctoral work at Stanford University.

MY laboratory has been engaged in cloning and characterizing mammalian genes that code for transcription factors. One focus is on signal transduction, specifically on mitogenic signaling, and another aims at identifying transcription factors that control kidney development. These two interests have recently intersected, as delineated below.

The Egr Family of Immediate-Early Transcription Factors

Extracellular "signals" in the form of neurotransmitters, growth factors, hormones, and matrix are known to control cellular phenotype. These agents lead to the generation of second messenger signals in the plasma membrane and cytosol. In turn, these biochemical events modulate the expression of so-called immediate-early genes (IEGs), whose induction does not require *de novo* protein synthesis. Several years ago, we and others identified several IEGs in the context of a mitogenic response, and more specifically in the transition of a cell out of a quiescent state (G_0) into G_1. Of particular interest to us has been a subset of IEGs that encode transcription factors (proteins that bind DNA and regulate gene transcription), since they might couple short-term responses in the form of second messenger events to long-term changes in gene expression instrumental in altering phenotype.

The best characterized of these immediate-early transcription factor genes include members of the *Fos* family. c-*Fos*, identified as the cellular homologue of the v-*Fos* oncogene present in two viruses that cause osteosarcomas, was discovered in 1984 to be an IEG in serum-stimulated fibroblasts. However, it was not until four years later that *Fos* was shown to be part of the AP-1 transcription factor complex, composed of Fos-Jun heterodimers and other less well characterized proteins. c-*Jun* was likewise first identified as the cellular homologue of the v-*Jun* transforming gene. It was suspected to be a transcription factor through homology to part of the yeast GCN4 protein.

In 1987 we discovered (concurrently with several other laboratories) the *Egr* family of IEGs. The best-characterized gene in this family is *Egr-1* (early growth response gene-1). *Egr-1* (also known as *Zif-268*, *Tis-8*, *NGFI-A*, and *Krox-24*) was isolated as a serum-inducible IEG in quiescent fibroblasts (G_0-G_1 transition), utilizing a differential screening protocol. The gene is induced by mitogen stimulation in every mammalian cell type tested, including B cells; T cells; kidney mesangial, glomerular, and tubular epithelial cells; hepatocytes; and vascular smooth muscle and endothelial cells. It is also induced by nerve growth factor in PC12 pheochromocytoma cells, a physiological context in which mitotic cells convert to a nonmitotic state.

The cDNA structure predicts a protein whose carboxyl terminus contains three zinc fingers of the Cys_2-His_2 type, first identified in the *Xenopus* transcription factor TFIIIA. This prediction has recently been verified by Carl Pabo (HHMI, Massachusetts Institute of Technology) and his colleagues through analysis of the crystal structure of the zinc finger domain cocrystallized with its target DNA sequence GCGGGGGCG.

We have been interested in identifying the events, from cell surface to nucleus, that modulate *Egr-1* expression. Although identification of such "upstream" or "proximal" events can either be attempted in the context of mitogenesis or in other situations in which *Egr-1* mRNA levels change, most of our work has been restricted to cell proliferation studies in fibroblasts. It has been found that multiple kinases regulate *Egr-1* expression. For example, activation of the PDGF (platelet-derived growth factor) receptor or the EGF (epidermal growth factor) receptor by their cognate ligands leads to *Egr-1* induction.

v-Src has been shown to regulate *Egr-1* independent of protein kinase C. More recently, we have found that v-Raf, a serine-threonine kinase, whose activation results from the convergence of diverse cell surface signals, leads to *Egr-1* induction. Furthermore, a dominant negative mutant of v-Raf will ablate the v-Src induction of *Egr-1*, suggesting that v-Src stimulates *Egr-1* via v-Raf. Even if *Egr-1* served as nothing more than a target

marker, studies like these are helping to define the circuitry among second messenger cascades.

Additional studies currently under way are aimed at defining structure-function relationships of the *Egr-1* protein. The protein contains several activation domains and a small modular repressor region whose sequence has been highly conserved in evolution. Nuclear localization is dependent upon a bipartite sequence, and DNA binding is determined by the three zinc fingers.

A major challenge ahead is to define a functional role for the *Egr-1* protein. Attempts at abrogating *Egr-1* activity may prove useful in defining a phenotype. In this regard, the use of antisense methodology—either as oligonucleotides or in the form of stable inducible vectors—is under investigation. The latter technique, when applied to c-*Fos*, has resulted in a remarkable and reversible inhibition of cell growth. Another approach is also available to inhibit *Egr-1* activity: nature has designed its own inhibitor of *Egr-1* function in the form of a tumor-suppressor gene, one whose absence is likely to lead to Wilms' tumor (see below).

Transcription Factors in Kidney Development

A second recent focus of our laboratory is on trying to define molecular events that characterize the development of the kidney. It is well known that metanephric mesenchymal (blastemal) cells convert into epithelial cells over a 4- to 5-day period (in the rat or mouse) in response to invasion of the ureteric bud. Little is known at the molecular and cellular level of the events that transpire during this conversion process. Our aim is to define a hierarchy of transcriptional regulators whose expression is modulated during nephrogenesis. Toward this end, several studies are under way.

One strategy is to identify zinc finger transcriptional regulators expressed in the kidney during differentiation. Using a so-called H/C-link probe directed against a region often conserved in zinc finger proteins, we have isolated from a kidney library a dozen cDNA clones encoding zinc finger proteins. These are now being characterized by their developmental profile (Northern blots and *in situ* hybridization) and by limited sequence analysis. One of these clones hybridizes at high stringency to the candidate Wilms' tumor antioncogene (*WT1*) identified recently by reverse genetic means in the laboratories of David Housman (Massachusetts Institute of Technology) and Gail Bruns (Harvard University). Another clone has identified a zinc finger protein that is expressed very early in human kidney development.

In a second approach, work is in progress to set up stem cell cultures from embryonic kidney for the development of cell lines arrested at different stages of differentiation. If these cells can be maintained and then made to differentiate in culture under appropriate conditions, they will be invaluable for studying the early molecular and cellular events in nephrogenesis.

Third, we are pursuing an interesting connection between Egr-1 and WT1. WT1 is a zinc finger protein; three of its four zinc fingers show a 60–70 percent similarity to the three zinc fingers of Egr-1. Indeed, WT1 and Egr-1 bind to a common sequence, as shown by the work of Frank Rauscher (Wistar Institute). We have recently shown (in collaboration with Dr. Rauscher) that WT1 acts as a transcriptional repressor, whereas Egr-1 is a transcriptional activator. These findings may have exciting biological implications, because several binding sites exist for the Egr-1/Wilms' tumor proteins in the promoter sequence of the insulin-like growth factor II (IGF-II).

IGF-II levels are known to be high in Wilms' tumor, and these levels fall during development. Thus the possibility exists that IGF-II is a target for the repressive action of the Wilms' tumor protein and for positive regulation by Egr-1. If this model holds, it will be a way to explain the findings of aberrant regulation of IGF-II in Wilms' tumor and will also provide major insight into the mesenchymal cell–to–epithelial cell conversion in the kidney.

Structure and Function of Voltage-Dependent Calcium Channels

Tsutomu Tanabe, Ph.D.—Assistant Investigator

Dr. Tanabe is also Assistant Professor of Cellular and Molecular Physiology at Yale University School of Medicine. He received his B.A., M.A., and Ph.D. degrees from Kyoto University in Kyoto, Japan. He began studies on the structure and function of the receptors and ion channels in excitable membrane when he was a graduate student in Shosaku Numa's laboratory. After he received his doctorate, he continued to conduct research in Dr. Numa's laboratory as a faculty member before coming to Yale.

VOLTAGE-dependent calcium channels play important roles in the regulation of a variety of cellular functions, including membrane excitability, muscle contraction, synaptic transmission, and secretion. At least four types of calcium channels have been distinguished by their electrophysiological and pharmacological properties.

Recently, molecular biological studies, combined with electrophysiology, have provided evidence that this diversity of calcium channels derives largely from differences in their pore-forming α_1-subunit and that the other subunits associated with α_1 can modify channel function. Furthermore, the diversity of the several subunits was shown to originate not only from differences in the genes encoding them but also from alternative splicing of their RNAs.

Several types of calcium channels are known to be coexpressed in single cells, and the cells apparently use them for different purposes. We are interested in the structure-function relationships of calcium channels and the molecular basis of one type's specialization.

Muscular dysgenesis (*mdg*) is a fatal autosomal recessive mutation of mice. It eliminates excitation-contraction (E-C) coupling and dihydropyridine (DHP)-sensitive calcium current of the slow L-type from skeletal muscle. Analysis of genomic DNA and skeletal muscle RNA indicates that the *mdg* mutation is associated with alterations of the structural gene for the skeletal muscle DHP receptor. Injection of an expression plasmid carrying the cDNA of the receptor restores both E-C coupling and skeletal L-type calcium current, suggesting a dual role of this molecule.

The restored coupling resembles that of normal skeletal muscle, which does not require entry of extracellular calcium ions. By contrast, injection into dysgenic myotubes of an expression plasmid carrying cDNA of the cardiac DHP receptor produces rapid, cardiac-like L-type current and cardiac-type E-C coupling, which does require calcium ion entry.

To investigate the molecular basis for these differences in calcium currents and E-C coupling, we expressed various cDNAs of the chimeric DHP receptor in dysgenic myotubes, with the following results. Expression of cDNAs encoding chimeras with regions of the skeletal muscle DHP receptor replacing one or more of the corresponding large, putative cytoplasmic regions of the cardiac DHP receptor showed that the region linking repeats II and III is a major determinant of skeletal muscle–type E-C coupling.

Expression of cDNAs encoding chimeras in which repeats of the skeletal muscle DHP receptor are replaced by corresponding repeats from the cardiac receptor showed that repeat I determines whether the chimeric calcium channel activation will be slow (skeletal muscle–like) or rapid (cardiac-like).

We are also focusing on the drug-binding sites of calcium channel molecules, including those of channel antagonist (L-type channel blocker), and the mechanism of calcium-dependent inactivation.

The Molecular Biology of Liver Regeneration

Rebecca A. Taub, M.D.—Associate Investigator

Dr. Taub is also Assistant Professor of Human Genetics at the University of Pennsylvania School of Medicine. She received her B.A. degree in biochemistry from Yale University, attended Yale University School of Medicine, and completed residency training in internal medicine at Yale. She was a postdoctoral fellow with Philip Leder at Harvard Medical School, where she later joined the Department of Internal Medicine as Assistant Professor before moving to Philadelphia.

THE liver has unusual properties of regeneration. When the two largest lobes of the liver are removed, leaving the smaller lobes completely intact, the remaining cells can grow and divide until the liver regains its former size, whereupon growth ceases. Many growth factors have been implicated in regulating this process, but the mechanisms remain poorly understood. After many years of study the same interesting questions about liver regeneration remain: What makes the liver start regenerating? What regulates liver growth during regeneration, allowing the liver to maintain its normal architecture? What makes the liver stop regenerating when it has regained its initial size?

Understanding liver regeneration will help explain how the liver responds to toxic damage or infections like hepatitis. Additionally, because increasing numbers of liver transplants are being performed and successful transplants require liver regeneration, it is essential to understand the biological and molecular bases for liver cell growth.

My colleagues and I are interested in determining what genes are expressed during liver regeneration and how their protein products are involved in regulating the process. It is important to determine if these genes are identical to those that regulate the growth of all cells. We are studying liver cell growth in two systems: a continuously growing liver cell line that responds to insulin as a growth factor and regenerating liver tissue from rats.

In our early studies, we identified more than 40 novel genes that are rapidly expressed in response to growth factors in liver cells. Although many of the genes are expressed in other growing cells and seem to be part of the general growth response, some of the genes are specific to growing liver. Many of these genes encode proteins that function in the cell nucleus, possibly regulating the cascade of gene expression occurring when cells grow. Some of these genes have altered expression in cancer cells, contributing to the aberrant proliferation.

Our studies have focused on understanding the actions of the proteins encoded by some of these genes and their potential roles in regulation of liver regeneration. Because we isolated so many novel genes, it was important to establish criteria for determining which of the encoded proteins are likely to have important regulatory roles in liver regeneration. We decided to explore further the exact roles of novel proteins falling into four functional categories.

The first category includes proteins that regulate the expression of genes. Because so many genes need to be turned on for liver regeneration to proceed, proteins that regulate gene expression are likely to be important.

The second category includes proteins that are secreted from cells and may function to regulate the growth of surrounding cells. The liver must maintain its cellular architecture during regeneration, and because it is made up of many different cell types, intercellular communications must exist during regeneration. Secreted proteins could be involved in mediating such intercellular signals.

The third category includes proteins whose expression is specific to regenerating liver and not other growing cells. We found several in this category, a few of which are highly expressed in regenerating liver. These proteins could have important functions in liver-specific growth.

The fourth category includes several genes whose expression is nicely induced in regenerating liver but abnormally high in the liver tumor cell line that grows in response to insulin. These genes could be functioning as oncogenes in the liver tumor cell line.

One of the novel genes that we isolated, *RL/IF-1* (regenerating liver inhibitory factor), encodes a protein that inhibits the activity of certain gene-transactivating proteins in the NF-κB/Rel family. We examined the relative activity of NF-κB/Rel proteins during liver regeneration and found to our surprise that although most of the proteins in this family remain in an inhibited or inactive form, one protein, PHF-1, becomes dramatically active in its ability to interact with gene sequences within minutes after hepatectomy. The activation of PHF-1 is the earliest change we

have observed in the liver posthepatectomy, and our findings suggest that PHF-1 could have a role as an initiating signal in liver regeneration.

LRF-1 (liver regeneration factor) is one of the proteins that we have studied that is highly expressed in regenerating liver and functions as a DNA-binding protein, controlling the expression of target genes. It falls into the category of so-called leucine zipper proteins, in which a stretch of amino acids containing evenly spaced leucine residues allows one molecule of LRF-1 to interact with another molecule of LRF-1 or of a related protein. LRF-1 activates or inhibits the expression of target genes in a singular manner, implying that LRF-1 has a unique role in regulating events in the regenerative process. As the relative level of LRF-1 complexes increases posthepatectomy, activation of the many liver-specific genes occurs, resulting in maintenance of normal liver function and metabolic homeostasis during regeneration.

Another goal of our research (supported in part by research funds from other agencies) is to understand the involvement of insulin in the regulation of hepatic growth. It is well known that diabetics have poor healing capacity, and early studies have shown that the livers of diabetic animals show poor regenerative capacity. Because we are studying a liver cell line that is growth-regulated by insulin, we have been able to compare the insulin-regulated growth response in these cells with the response during liver regeneration. Differences in expression of more than 10 of the novel genes we have identified suggest that insulin, if it is an important growth factor during liver regeneration, must act several hours after the hepatectomy and is not an initiating factor. Additionally, in examining the encoded function of several genes that are aberrantly regulated in the hepatic cell line, we are learning more about the specific signaling pathways involved in hepatic growth.

Protein-Tyrosine Phosphatases and the Control of Lymphocyte Activation

Matthew L. Thomas, Ph.D.—Assistant Investigator

Dr. Thomas is also Associate Professor of Pathology and Assistant Professor of Molecular Microbiology at Washington University School of Medicine, St. Louis. He received his Ph.D. degree from the University of Utah after completing his thesis research at Harvard Medical School. His postdoctoral training was done with Alan Williams at Oxford University and with Ian Trowbridge at the Salk Institute. Prior to his present appointment, Dr. Thomas was an Established Investigator of the American Heart Association.

THE molecular mechanisms by which cells respond to their environment are a central theme in many areas of medical research. Our interests are in understanding how signals received by proteins of leukocyte surface membranes result in changes in a wide variety of cellular processes. In particular, we are interested in lymphocyte activation.

Lymphocytes, by virtue of their ability to recognize an infinite variety of foreign antigens, play a key role in effecting and regulating an immune response. Upon binding antigen, lymphocytes undergo profound biochemical changes to engage the cellular machinery required for clonal expansion and to produce molecules needed to fight infection. A key regulatory mechanism by which many different cell types control the signal transduction process is phosphorylation and/or dephosphorylation of distinct protein-tyrosine residues of specific substrates. This process is controlled by families of enzymes that either add phosphate (protein-tyrosine kinases) or remove phosphate (protein-tyrosine phosphatases). The research interests of my laboratory are centered on how protein-tyrosine phosphatases control lymphocyte activation.

The protein-tyrosine phosphatase family can be divided into two main branches: transmembrane and intracellular. The leukocyte-common antigen CD45 is a major transmembrane protein-tyrosine phosphatase of lymphocytes and is expressed by all nucleated cells of hematopoietic origin. To study the function of this molecule, we generated nontransformed T cell clones deficient in the expression of CD45. The deficiency resulted in the cells' inability to respond to antigen. CD45-deficient cells still respond to other proliferative signals, such as stimulation by growth factors or plant lectin mitogens or direct activation of protein kinase C. However, these cells cannot proliferate or produce cytokines in response to T cell antigen receptor stimulus, and their ability to cytolyze target cells is also impaired. Therefore the protein-tyrosine phosphatase CD45 is required for antigen-induced T cell activation.

We are investigating the molecular mechanism of antigen-induced CD45-controlled activation by examining the differences in tyrosine phosphorylation among the CD45-expressing and -nonexpressing cell lines. We have observed increased tyrosine phosphorylation of members of the Src-tyrosine kinase family of proteins in the CD45-deficient cells, indicating that they are potential substrates for CD45. Three members of the Src family, $p56^{lck}$, $p59^{fyn}$, and $p62^{yes}$, are expressed by T cells, and recent studies have indicated that kinase activity for all three is decreased in the CD45-deficient T cells. Src family members are negatively regulated by phosphorylation of a carboxyl-terminal tyrosine residue. CD45 may serve to dephosphorylate the carboxyl-terminal tyrosine site and thus function to activate members of the Src family. The inability of antigen to activate CD45-deficient T cells implies that activation of Src family members is critical to antigen-induced T cell activation and that CD45 is the phosphatase important in initiating this process. This work is supported in part by a grant from the National Institutes of Health.

Immortalization of CD45-deficient T cells by fusion with a transformed thymoma cell line has permitted the development of a model system in which function can be reconstituted by cDNA transfection. We are currently analyzing regions of functional importance in the CD45 molecule by reconstituting the cells with altered CD45 cDNAs. These experiments should allow precise definition of the regions of the molecule important in regulating T cell activation.

To analyze further how protein-tyrosine phosphatases control lymphocyte activation, we have isolated multiple cDNAs that encode protein-tyrosine phosphatases expressed by leukocytes. LRP is a transmembrane phosphatase whose exterior domain is predicted to be a highly glycosylated, elongated rod. This type of structure is also found at the amino terminus of CD45's exterior domain and may provide a novel means of regulating the phosphatase activity. To understand its function and regulation, we have developed a monoclonal antibody that recognizes the native protein. Use of this antibody has allowed us to

examine the expression of the protein by various leukocyte populations, the extent of post-translational modification, and the ability of the antibody to induce changes in lymphocyte activation and proliferation.

PEP and SHP are both intracellular protein-tyrosine phosphatases expressed primarily by leukocytes. PEP contains a large carboxyl-terminal extension following the phosphatase domain. This region is unusual when compared with other intracellular phosphatases, in that it contains sequences indicative of nuclear localization and rapid protein turnover. By tagging the protein with a sequence that can be traced, we have obtained preliminary results indicating that PEP does indeed localize to the nucleus. It is possible, therefore, that PEP is involved in regulating gene transcription or other nuclear functions.

SHP contains two *src*-homology region-2 (SH2) domains linked in tandem immediately amino terminal to the PTPase catalytic domain. SH2 domains are found in proteins involved in transducing mitogenic signals, such as all the nonreceptor protein-tyrosine kinases, the $p21^{ras}$ GTPase-activating protein, and the γ-isoform of the phosphatidylinositol-specific phospholipase C. Functionally, SH2 domains bind phosphorylated tyrosine residues. Thus a protein that contains SH2 domains serves to amplify and direct the biochemical pathway of a response that is induced by an initial increase in tyrosine phosphorylation. Recent results indicate that the SHP SH2 domains are capable of binding a low-molecular-weight phosphotyrosine protein. The existence of a phosphatase that contains SH2 domains suggests that SHP may serve to modulate kinase-induced signals. We are examining the plausibility of this model by identifying the specific proteins with which SHP interacts.

Through an understanding of how protein-tyrosine phosphatases effect lymphocyte activation, we hope to gain a more thorough knowledge of the biochemical steps involved in controlling and regulating an immune response. The functional characterization of multiple protein-tyrosine phosphatases expressed by lymphocytes has allowed us to define steps in which members of this family function in lymphocyte activation.

Molecular Regulation of Lymphoid Cell Growth and Development

Craig B. Thompson, M.D.—Associate Investigator

Dr. Thompson is also Associate Professor of Internal Medicine and of Microbiology and Immunology at the University of Michigan Medical School. He received his undergraduate degree from Dartmouth College and his medical degree from the University of Pennsylvania. Following an internship and residency at the Peter Bent Brigham Hospital, Boston, he spent eight years as a research medical officer in the United States Navy. During this time, he conducted his research at Boston University, the Uniformed Services University of the Health Sciences, the Fred Hutchinson Cancer Research Center, and the Naval Medical Research Institute.

THE lymphoid immune system comprises two major cell types: the T cell, which identifies and destroys cells expressing foreign proteins; and the B cell, which secretes antibodies that bind to foreign substances, targeting them for elimination. The central role of the system in the natural resistance to infectious diseases is demonstrated by the infections encountered by patients with immunologic deficiencies. Many of the serious infectious and neoplastic complications associated with the acquired immune deficiency syndrome (AIDS) are the result of depletion of the helper T cells.

A better understanding of the molecular mechanisms associated with generation of both B and T cells during development would aid in our ability to understand and treat various immunodeficiencies. The goals of our laboratory are to understand the molecular events associated with the development of the lymphoid immune system and to define the mechanisms by which the functions of these cells are controlled.

T cells are divided into two major subsets: helper T cells, which produce the lymphokines that regulate immune responses, and cytotoxic T cells, which can kill cells expressing foreign proteins. The cooperation of these two cell types is needed for the immune system to reject a foreign cell or a cell bearing foreign proteins. Initiation of a helper T cell response requires activation of the T cell receptor by foreign histocompatibility genes. Recent evidence suggests, however, that T cell receptor activation alone does not lead the helper cell to produce sufficient amounts of lymphokines to initiate an immune response. Additional co-stimulatory signals are required.

Previously we showed that the CD28 receptor expressed on helper T cells serves as a surface component of a signal transduction pathway that can enhance T cell lymphokine production. *In vitro*, interaction of CD28 with its natural ligand B7, expressed on activated B cells or macrophages, can act as a co-stimulus to such enhancement. Now we have evidence that CD28 activation of primary T cells is also a required co-stimulatory event in the initiation of a cell-mediated immune response.

We have investigated the role of the CD28 pathway in allogenic responses by using a soluble CD28 receptor homologue termed CTLA-4-Ig, produced by our collaborator Peter Linsley. This protein is a recombinant that displays an affinity for the CD28 ligand, B7, approximately 20-fold higher than does cell-bound CD28. Therefore it acts as a competitive inhibitor of CD28 engagement. *In vitro*, CTLA-4-Ig was found to be able to inhibit completely the ability of cells from one rat strain to respond against cells from another.

Based on these data, the ability of CTLA-4-Ig to prevent the induction of organ graft rejection was tested in a cardiac transplant model. Transplanted hearts in control animals receiving no immunotherapy are rejected within one week of transplantation. In contrast, when activation of the CD28 pathway is blocked by daily administration of CTLA-4-Ig, rejection of the cardiac transplant is prevented.

These data support the hypothesis that helper T cell co-stimulation by the CD28-ligand, B7, is a required event in the initiation of a T cell–mediated immune response. Furthermore, the work suggests that preventing B7 activation of a T cell may not only help to prevent transplant rejection but may also help to decrease the severity of autoimmune diseases such as rheumatoid arthritis. Work to address this issue is being planned.

The major role of the B cell immune system is to generate the approximately 10 million different antibody molecules needed to protect the body from foreign substances. B cells derive their name from the bursa of Fabricius, a developmental organ in birds that is required for B cell maturation. Mammals lack this organ but are still able to generate a B cell immune system. Over the past several years, our laboratory has investigated the role of the bursa in B cell development in the chicken. We are attempting to characterize dif-

ferences in the generation of B cells in mammals and birds to shed light on mammalian B cell production.

Our studies have helped to demonstrate that the primary molecular mechanism by which chickens generate a wide variety of antibody molecules is different from that used by mammals. Antibody diversity in the chicken is achieved by a process known as gene conversion, which occurs during B cell development in the bursa of Fabricius. Gene conversion requires that the immature B cell migrate to the bursa, and that the mature B cell, having undergone conversion, emigrate to the peripheral lymphoid organs. Understanding how this migration is controlled in a regulated fashion should lead to insights into how cells in a multicellular organism migrate during development.

The Molecular Basis of Metamorphosis

Carl S. Thummel, Ph.D.—Assistant Investigator

Dr. Thummel is also Assistant Professor of Human Genetics at the University of Utah School of Medicine. He obtained his undergraduate degree in biology from Colgate University and his Ph.D. degree in biochemistry, working with Robert Tjian, at the University of California, Berkeley. He received postdoctoral training in the laboratory of David Hogness at Stanford University.

THE fruit fly *Drosophila melanogaster* provides an ideal model system for studying the development of eukaryotes. Three-quarters of a century of biological, physiological, and genetic experiments, combined with recent intensive molecular studies, has led to a greater understanding of its development than that of any other higher organism.

Halfway through the fly's life cycle, a pulse of the steroid hormone ecdysone triggers a dramatic morphological transformation, from the relatively immobile feeding larva to a highly motile, reproductively active adult fly. We are studying the molecular basis of the ecdysone-induced regulatory mechanisms that allow metamorphosis to proceed.

When the larva begins to undergo metamorphosis, its salivary glands contain giant polytene chromosomes that can be visualized by light microscopy. These 500-fold overreplicated, interphase chromosomes lie in register beside one another. A characteristic banding pattern along the length of the polytene chromosomes allows any gene of interest to be located precisely. Regions of the genome that are undergoing transcription are often represented by large areas of decondensed chromatin that can be seen as puffs. Thus the transcriptional activity of specific genes at specific times can be followed in these chromosomes by observing the appearance and disappearance of puffs during development.

Approximately 10 puffs can be distinguished when the salivary gland chromosomes first become large enough to see. These puffs remain until the end of the larval phase, when the burst of ecdysone triggers a dramatic change in the puffing pattern. Approximately six puffs are induced directly by the steroid hormone. These early ecdysone-inducible puffs appear to encode regulatory proteins that repress their own expression and induce the formation of over 100 late puffs. This second wave of puffs is believed to encode the proteins responsible for initiating metamorphosis.

By isolating and characterizing the ecdysone-inducible genes that lie within the early puffs, we hope to learn how these genes are induced by the hormone and how their encoded proteins might function in a regulatory capacity. In a broader sense, this project provides a model system for characterizing the role of steroid hormones in regulating gene expression and for addressing the question of how gene hierarchies are controlled during development.

Our studies have focused on *E74*, an ecdysone-inducible gene that is located within the large early puff at position 74EF in the polytene chromosomes. This unusually complex gene encodes three nested mRNAs that derive from unique start sites but share a common 3′ end. The distal promoter directs the synthesis of a 60-kb primary transcript that is spliced to form the 6-kb *E74A* mRNA. Two other promoters, located 40 kb downstream from the *E74A* promoter, direct the synthesis of 4.8- and 5.1-kb *E74B* mRNAs. Although the *E74A* and *E74B* mRNAs are distinct from one another by virtue of their unique 5′ exons, the majority of these mRNAs are identical, derived from a common set of three 3′ exons. This nested arrangement of the *E74* transcripts leads to the synthesis of two related *E74* proteins that have unique amino-terminal domains joined to a common carboxyl-terminal domain.

The sequence of the carboxyl terminus of the *E74* proteins is very similar to a portion of the protein encoded by the *ets* oncogene. This 85–amino acid ETS domain defines a family of proteins and has been shown to function as a site-specific DNA-binding domain that recognizes a purine-rich DNA sequence. Studies of proteins related to oncogenes, such as *E74*, may help us learn more about how the normal counterparts of these disease genes function during development.

By using antibody detection techniques to localize the E74A protein bound to the giant polytene chromosomes, we have identified approximately 70 binding sites, most of which correspond to late ecdysone-inducible puffs. Based on this observation, we predict that at least one function for the E74A protein is to activate late gene expression. In support of this prediction, many late puffs are either reduced or absent

in the polytene chromosomes of mutant animals that do not express E74A protein. Our long-term goal is to identify some of these late genes in order to determine what role they might play during metamorphosis and whether they are directly regulated by E74A.

The precise timing of the genetic response to ecdysone can be clearly seen in the pattern of puffs that arise at the onset of metamorphosis. Our characterization of *E74* transcription has provided insights into how timing can be built into a genetic regulatory hierarchy. Ecdysone directly activates the *E74A* promoter, resulting in a dramatic induction of the 6-kb mRNA. This transcript, however, does not appear in the cytoplasm until one hour after promoter activation. The delay corresponds quite closely to the time it takes for RNA polymerase to traverse the 60-kb transcription unit, indicating that the length of the *E74A* unit functions as a timer to delay significantly the appearance of the encoded mRNA. The unusual length of the *E74A* primary transcript sets it apart from most transcription units in *Drosophila*, which are only slightly longer than the final processed mRNA.

Ecdysone also directly activates the 20-kb *E74B* transcription unit. In agreement with its primary transcript length, mature *E74B* mRNA appears between 15 and 30 minutes after ecdysone addition. Thus the structure of the *E74* gene dictates an invariant order of appearance of its transcripts in response to ecdysone.

The earlier appearance of *E74B* mRNA is enhanced by its activation at an approximately 25-fold lower ecdysone concentration than *E74A*. *E74B* is further distinguished from *E74A* by its repression at a significantly higher ecdysone concentration than that required for its induction, close to the concentration required for *E74A* activation. These regulatory properties lead to an ecdysone-induced switch in *E74* expression, with an initial burst of *E74B* transcription followed by a burst of *E74A* transcription. These studies provide a means to translate the profile of a steroid hormone pulse into different amounts and times of regulatory gene expression that, in turn, could direct different developmental responses in a temporally and spatially regulated manner.

To date, *E74* has provided a valuable paradigm for our molecular characterization of the ecdysone regulatory hierarchy. We have recently begun to extend our studies to include three other early genes, all of which, like *E74*, are unusually long and encode multiple DNA-binding proteins. Characterization of the temporal regulation of these early genes, like that described above for *E74*, has confirmed that both promoter sensitivity to ecdysone and primary transcript length contribute to the timing of early gene activation.

In addition, these studies have allowed us to divide the early transcription units into two classes. One class of transcripts is activated by a low ecdysone concentration, like *E74B*, and appears to play earlier roles in the regulatory hierarchy. This class includes transcripts encoding the ecdysone receptor, needed for initiating the genetic response to the hormone. The second class, typified by *E74A*, is activated by a higher hormone concentration and appears to play a later role in the hierarchy.

In addition, analysis of the effects of early gene mutations on both early and late gene expression should provide valuable clues regarding functional regulatory interactions within the hierarchy. The predictions that arise from these genetic studies can then be tested at the molecular level. An additional long-term goal will be to isolate more early and late ecdysone-inducible genes to extend our understanding of this complex developmental process.

The Regulation of Mammalian Development

Shirley M. Tilghman, Ph.D.—Investigator

Dr. Tilghman is also Howard A. Prior Professor of the Life Sciences in the Molecular Biology Department at Princeton University and Adjunct Professor of Biochemistry at the University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School. She obtained a B.Sc. degree at Queen's University in Kingston, Ontario, Canada. Following two years in Sierra Leone, West Africa, where she taught secondary school, she attended graduate school at Temple University in Philadelphia, where she received her Ph.D. degree in biochemistry. Her postdoctoral work was done with Philip Leder at NIH. Before joining the faculty at Princeton, Dr. Tilghman held positions at Temple University and the Institute for Cancer Research, Philadelphia.

ORDERLY development of the mammalian embryo requires the appropriate activation and subsequent modulation of genes in a spatial and temporal manner. For the vast majority of genes, both the mother's and father's copies are activated and modulated identically, but for a small class of genes, only the mother's or father's copy is expressed. Such genes are parentally imprinted. That is, during the process that generates eggs or sperm, these genes are marked in such a way that the resulting embryo can distinguish the parental origin and express it accordingly. Our laboratory is studying a locus that encodes at least two imprinted genes on the distal end of mouse chromosome 7.

One, the insulin-like growth factor II gene (*Igf2*) encodes a fetal-specific growth factor that is exclusively expressed from the paternal chromosome. The other gene is *H19*, which encodes an RNA, evolutionarily conserved, that is only found in high abundance during fetal development in tissues originating in endoderm and mesoderm. Unlike *Igf2*, *H19* is exclusively expressed from the maternal chromosome. These genes lie in tandem about 75 kilobases (kb) of DNA apart and are expressed in a very similar manner during mouse embryogenesis.

We are investigating the activation and role of these two differentially imprinted genes. The function of the H19 RNA is unknown. Its pattern of sequence conservation in mammals is reminiscent of other functional RNAs, such as those associated with telomerases and RNase P. We are using both genetic and biochemical approaches to understand its role during development.

Over the past 50 years a large number of mutations have been described that affect all aspects of mouse physiology. However, molecular access to the genes, which would allow us to identify those of developmental importance, has been difficult because the mouse genome is so large. Advances in DNA analysis and cloning methods have effectively reduced the barriers to studying these genes at the molecular level. We have generated a yeast artificial chromosome library that now contains 2.5 copies of the mouse genome in over 28,000 yeast strains, each of which harbors a segment of mouse genomic DNA averaging 275 kb. The library has been constructed to serve the mouse genome community, and over 60 laboratories worldwide have so far screened it successfully. This work is supported by a grant from the National Institutes of Health.

We have used the mouse *Fused* (*Fu*) locus to test the utility of the library for isolating large chromosomal DNA segments. *Fu*, a dominant mutation on mouse chromosome 17, generates kinky tails in heterozygotes and early embryonic lethality in homozygotes. The lethality is associated with overgrowth of neuroectoderm and duplications of the body axis. To localize *Fu* precisely, 1,000 progeny were generated by backcrossing mice carrying the *Kinky* allele of *Fu* with a distantly related wild mouse, *Mus spretus*. Because these mice are genetically very distinct, it is easy to follow the segregation of genes in their progeny.

The offspring were scored for seven molecular markers that map in a small interval around *Fu*, and a high-density genetic map gives us molecular landmarks every 100–200 kb of DNA. One marker, a pseudogene of the α-globin gene family, cannot be separated from *Fu* in this cross, suggesting that it is very close to the gene—within 100–200 kb. We have used this close proximity to isolate approximately 650 kb of DNA around the marker. By comparing the segregation of this DNA with that of *Fu*, we should be able to pinpoint *Fu*'s location.

A similar approach has been adopted for the *piebald* (*s*) locus on mouse chromosome 14. Mice carrying the original *s* mutation have spotted coats, a result of the absence of melanocytes in genetically specified regions of the midsection. Mice carrying more-severe *s* mutations are almost entirely white. In addition, they develop

megacolon, as a result of the absence of enteric ganglia in the gut. Both melanocytes and enteric ganglia develop from the neural crest, a migrating cell population that contributes to many cell types. By studying the *s* gene, we hope to gain insight into the genes that control the birth, migration, and maturation of these cells.

A high-density map around *s* is under way, using mice carrying the original *s* mutation and another wild strain, *Mus castaneus*. In addition, the molecular studies are greatly aided by the many different mutant alleles of *s* created over the past 40 years at the Oak Ridge National Laboratory by William and Leane Russell. These alleles, many of which are deletions, will be invaluable in mapping the region on chromosome 14 that contains the *s* gene.

To gain insight into the nature of the *s* mutation, we are also comparing the behavior of neural crest cells in *s*-bearing and normal mice. For this purpose we have exploited a transgenic mouse strain carrying a β-galactosidase gene that is expressed in early neural crest cells and their derivatives. To study later stages in melanocyte development, we are utilizing an antibody to the cell surface protein c-kit, which is expressed in these cells. We will thus identify the stage in neural crest development that the *s* mutation affects.

Mechanisms of Gene Regulation in Animal Cells

Robert Tjian, Ph.D.—Investigator

Dr. Tjian is also Professor of Molecular and Cell Biology at the University of California, Berkeley, and Adjunct Professor of Biochemistry and Biophysics at the University of California, San Francisco. He received the Ph.D. degree in biochemistry and molecular biology from Harvard University. Following this he was a Junior Fellow of the Harvard Society of Fellows while a resident at Cold Spring Harbor Laboratory and later a staff investigator there before moving to Berkeley. His honors include the Monsanto Molecular Biology Award of the National Academy of Sciences and the Pfizer Award for Enzymology. Dr. Tjian was recently elected to the National Academy of Sciences.

THE main research interest of our laboratory is the mechanism by which the genetic information stored in DNA molecules is retrieved in a controlled and orderly fashion during the process called transcription, which leads to the production of specific proteins in animal cells. We have taken a biochemical approach to the problem of gene control and have devised various means of isolating the cellular components responsible for transcription and of reconstructing this complex reaction in the test tube. In this way we are studying how specific genes are turned on and off during cell growth and development of eukaryotic organisms. The mechanisms that govern the activation of genes are of fundamental importance in understanding the normal metabolic processes that maintain and perpetuate living cells, as well as in deciphering cellular and genetic disorders.

Biochemical Analysis of Cancer Genes

A living cell contains hundreds of thousands of protein molecules, each carrying out its allotted function. However, if either the production or action of these molecules is altered, severe malfunction can result, such as uncontrolled growth leading to cancer. Thus certain key molecules in the milieu of a normal cell's constituents have the potential to cause tumors when their function is disrupted. Such molecules are called oncoproteins.

Our group previously isolated from human cells a family of rare proteins that have subsequently been shown to be oncogenic, encoded by the *jun* and *fos* genes. These regulatory proteins are normally involved in controlling the action of many other genes, but when their activities are perverted—for example, by viruses—they can lead to the production of cancer-causing cells.

Recent advances in the study of the nuclear oncogene *jun* reveal that the ability of its protein to activate transcription is regulated by a cell-type–specific inhibitor that interacts with a unique portion of the molecule, rendering it less potent. It is anticipated that the isolation and characterization of this specific negative regulator of *jun* will help unravel the molecular signaling pathways responsible for transducing information from the outside of the cell into the nucleus, where gene expression is controlled. Moreover, such inhibitors of *jun* activity may also be found to represent new members of the anti-oncogenic or tumor-suppressor family of biological regulatory molecules.

Studies of Trans-activating Proteins That Regulate Gene Expression

A major hurdle has been the development of biochemical techniques that allow the purification of certain rare and fragile regulatory proteins. Through use of DNA-affinity chromatography procedures pioneered in this laboratory, it is now possible to isolate such transcription proteins and, in turn, to clone molecularly the genes that encode them. The ability to proliferate this biologically important class of genes provides a powerful approach toward understanding their structure and function. In the past two years the laboratory has isolated and characterized some 10 different genes that are directly responsible for the tissue-selective, temporally programmed, and basal-level control of gene expression in animal cells.

In addition, these recent studies are beginning to reveal new concepts regarding the surprisingly modular construction of the derivative proteins as well as their unusual plasticity and functional flexibility. Most importantly, specific structural motifs that lie within these proteins have been recognized as carrying out distinct functions. These findings provide the theoretical basis for analysis of other as yet undiscovered regulatory factors and will greatly aid our ability to decipher their mechanisms of action.

How Promoter-Specific Regulators Trigger Transcription

One of the remaining fundamental mysteries is the mode of action by which sequence-specific

DNA-binding proteins, such as the prototype human factor Sp1, direct transcriptional interactions. To address this critical issue, our group recently fractionated and isolated the multiple components necessary to reconstitute transcription. In the process of dissecting the general transcriptional apparatus, we discovered two previously undetected components that serve as the functional bridge between upstream trans-activators and the initiation complex. These novel factors appear to be part of the missing link that directs promoter-selective transcription in animal cells, and it is likely that they will be members of a diverse and essential class of regulatory proteins.

Indeed, this past year has seen significant progress in the purification and characterization of transcriptional "coactivators" and TATA-binding protein (TBP)-associated factors (TAFs). In particular, several TAFs and coactivators have been purified from both *Drosophila* and human cells. In addition, the genes encoding this novel class of transcription factors have recently been cloned. The structure and function of these multisubunit complexes should prove to be very revealing.

One of the most exciting and unexpected findings this year was the discovery that the TBP, a general transcription factor thought to be only responsible for RNA polymerase II transcription of mRNA, is also an integral subunit of a TBP-TAF complex responsible for recognition of the RNA polymerase I promoter. Most interestingly, the RNA polymerase I complex carries out all the functions ascribed to the species-specific transcription factor SL1, and the subunit composition reveals the presence of TBP and three novel TAFs, apparently adapted uniquely to direct RNA polymerase I transcription. This surprising finding provides a unifying mechanism of transcription initiation.

Transcription of Developmentally Regulated Genes

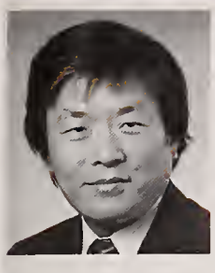
One of our long-term interests is the mechanisms underlying regulation and expression in the development of higher organisms. We have begun to address this issue in two ways: first, by initiating a series of *in vitro* experiments aimed at dissecting the transcriptional regulation of *Drosophila* genes, including the alcohol dehydrogenase, *Ultrabithorax*, *Antennapedia*, dopa decarboxylase, and *hunchback* genes. A major advance was the development of *in vitro* transcription reactions from staged *Drosophila* embryos that accurately initiate RNA synthesis and recapitulate the temporal program of transcription displayed by these tissue-specific and developmentally regulated genes.

A second approach has been to investigate the regulatory mechanism of RNA polymerase initiation factors *in vivo*. Various systems have been adapted to introduce altered genes back into cells or whole organisms to study their patterns of expression.

These *in vitro* and *in vivo* studies have recently led to two exciting results. First, a negative regulator of a developmentally important gene has been identified by direct biochemical means, and its mode of operation can now be dissected. The finding of specific transcriptional repressors is of particular importance because it is thought that an interplay of positive activators and negative regulators is seminal to the spatially restricted patterns of expression observed during embryogenesis.

Also arising from *Drosophila* studies are the identification and subsequent biochemical characterization of a transcription factor that appears to govern the expression of genes in cells of the central nervous system. The gene encoding this neurogenic-specific activator has recently been isolated, and its structure is expected to reveal interesting information. Advantages in the use of fruit flies include the ability to probe the developmental and tissue-specific function of this neurogenic regulator in a rapid and highly informative manner not readily applicable to mammalian cells. These studies are expected to yield new insights concerning the tissue-specific distribution and temporal timing of expression during development.

Studies on T Lymphocytes and Mammalian Memory



Susumu Tonegawa, Ph.D.—Investigator

Dr. Tonegawa is also Professor of Biology at the Massachusetts Institute of Technology. He received a B.S. degree in chemistry from Kyoto University in Kyoto, Japan, and a Ph.D. degree in biology from the University of California, San Diego. His postgraduate training and research were at UCSD in the laboratory of Masaki Hayashi and at the Salk Institute with Renato Dulbecco. Dr. Tonegawa was a member of the Basel Institute for Immunology in Basel, Switzerland, before joining the Department of Biology and Center for Cancer Research at MIT. He was awarded the Nobel Prize for physiology or medicine in 1987.

T lymphocytes play a pivotal role in the body's defense against a variety of infectious agents and malignant tumors as well as in the rejection of grafted foreign tissues. Our laboratory continues to study the development and functions of T lymphocytes, with a particular focus on more recently discovered and less-characterized $\gamma\delta$ T cells. In addition, our work on mammalian memory has progressed significantly during the past year.

Our study on $\gamma\delta$ T cells could be divided into three categories: development and selection, specificities, and function. As for the early development of these cells, one major issue that we continued to pursue is the mechanism by which the $\gamma\delta$ lineage segregates from the $\alpha\beta$ lineage from common progenitor cells. Following our earlier observations that differential activation of the transcriptional silencer associated with the γ genes of the T cell receptor (TCR) plays a pivotal role in the cell lineage segregation, we characterized the silencer DNA elements in detail and identified proteins that bind to these elements. We intend to clone the genes encoding the silencer-binding proteins and produce mice with mutations in these genes. Analysis of the mice should be highly informative in the dissection of the lineage segregation mechanism.

$\alpha\beta$ T cells are known to undergo in the thymus a critical maturation step called positive selection. This depends on an appropriate interaction between the TCRs and a self determinant encoded in part by the genes of the major histocompatibility complex (MHC). Following our earlier study indicating that epithelium-associated $\gamma\delta$ T cells (called s-IEL and vut-IEL) undergo a similar positive selection, we have now obtained evidence of positive selection for circulating $\gamma\delta$ T cells. Thus the development of a transgenic $\gamma\delta$ T cell clone (KN6) specific for an MHC class I product (encoded by the *T22*^{b} gene) was shown to be blocked at an immature state in a mouse genetically deficient in the expression of MHC class I gene products.

We also studied T cell development by exploiting the embryonic stem (ES) cell gene-targeting method. We produced four types of mutant mice relevant for the analysis of T cell development and functions. The first and second types of mice have mutations in the TCRα and TCRβ genes, respectively, and are deficient in mature $\alpha\beta$ T cells but not in mature $\gamma\delta$ T cells. The third type of mouse has a TCRδ gene mutation and is deficient in $\gamma\delta$ T cells but not in $\alpha\beta$ T cells. The fourth type was produced by mutating the *RAG-1* gene, whose product is required for somatic rearrangement of immunoglobulin and TCR genes. Neither mature T cells nor B cells are present in the *RAG-1* mutant mice.

The initial analysis of these mice indicated that there is no major interaction between the development of $\alpha\beta$ and $\gamma\delta$ T cells and that the mutations in the various genes block lymphocyte development at distinct stages. Detailed analysis of the immature lymphocytes accumulating in these mutant mice is expected to augment our understanding of lymphocyte development.

Specificities of $\gamma\delta$ T cells were studied by characterizing the self-reactive $\gamma\delta$ T cell subset (V_4C_1) on the one hand, and by testing our hypothesis that TL class I molecules have evolved to present peptides to $\gamma\delta$ TCRs. In the V_4C_1 subset study, we produced three types of monoclonal antibody (mAb), each of which blocks self-reactivity of these TCRs *in vitro*. These mAbs recognize 1) a determinant specifically present on the $\gamma\delta$ TCR utilized as the immunogen (called clonotypic), 2) determinants present on all $V\gamma_4C\gamma_1$-expressing TCR, or 3) determinants present on the receptor for vitronectin, a family of cell adhesion molecules previously implicated as coreceptors for this T cell subset. These mAbs will be extremely useful in the elucidation of the ligand for the self-reactive $\gamma\delta$ T cell subset, as well as for the analysis of its development and tissue localization.

We are studying the general peptide-presenting role of TL class I molecules by producing mouse mutants lacking a cluster of these molecules. We have thus far identified ES cell clones

devoid of a large number of TL class I genes. Also being pursued is biochemical characterization of a TL class I molecule encoded by the gene $T3^b$, whose product has been shown to be expressed specifically on the surface of gut epithelial cells with which a $\gamma\delta$ T cell subset, i-IEL, is associated. We intend to purify the putative peptides bound to the $T3^b$ molecule and subject them to microsequencing. The sequences may lead to the identification of the antigenic proteins.

We have pursued functions of $\gamma\delta$ T cells by immunizing KN6 transgenic mice with the ligand-bearing C57BL/6J spleen cells and following the proliferation of the transgenic cells and the disappearance of the injected spleen cells from the host spleen. These studies indicate that KN6 $\gamma\delta$ T cells are capable of responding to the allogenic cells by cytotoxicity.

Functions of $\gamma\delta$ T cells have also been studied by following the antibody response of $\alpha\beta$ T–deficient mutant mice to thymus-dependent and -independent antigens. So far these mice respond only to the thymus-independent antigen. A possible role of $\gamma\delta$ T cells in this type of antibody response is being studied using $\gamma\delta$ T–deficient mutant mice. The $\alpha\beta$ T– or $\gamma\delta$ T–deficient mice have also been infected with listeria, mycobacteria, or malaria. Resistance and/or immunity against these infectious agents will be determined. We are also studying the response of the mutant mice to skin grafts and injected tumor cells.

In addition to the studies on $\gamma\delta$ T cells, we found, using the multigene transfection technique, that the $\alpha\beta$ TCR-CD3 complex can be expressed on the surface of nonlymphoid cells without either the CD3γ or CD3δ subunit. This suggests the intriguing possibility that these multiple forms of TCR-CD3 complex are utilized by normal T cells for differential purposes. The hypothesis is being tested by producing mouse mutants that are defective either in the CD3γ or CD3δ gene.

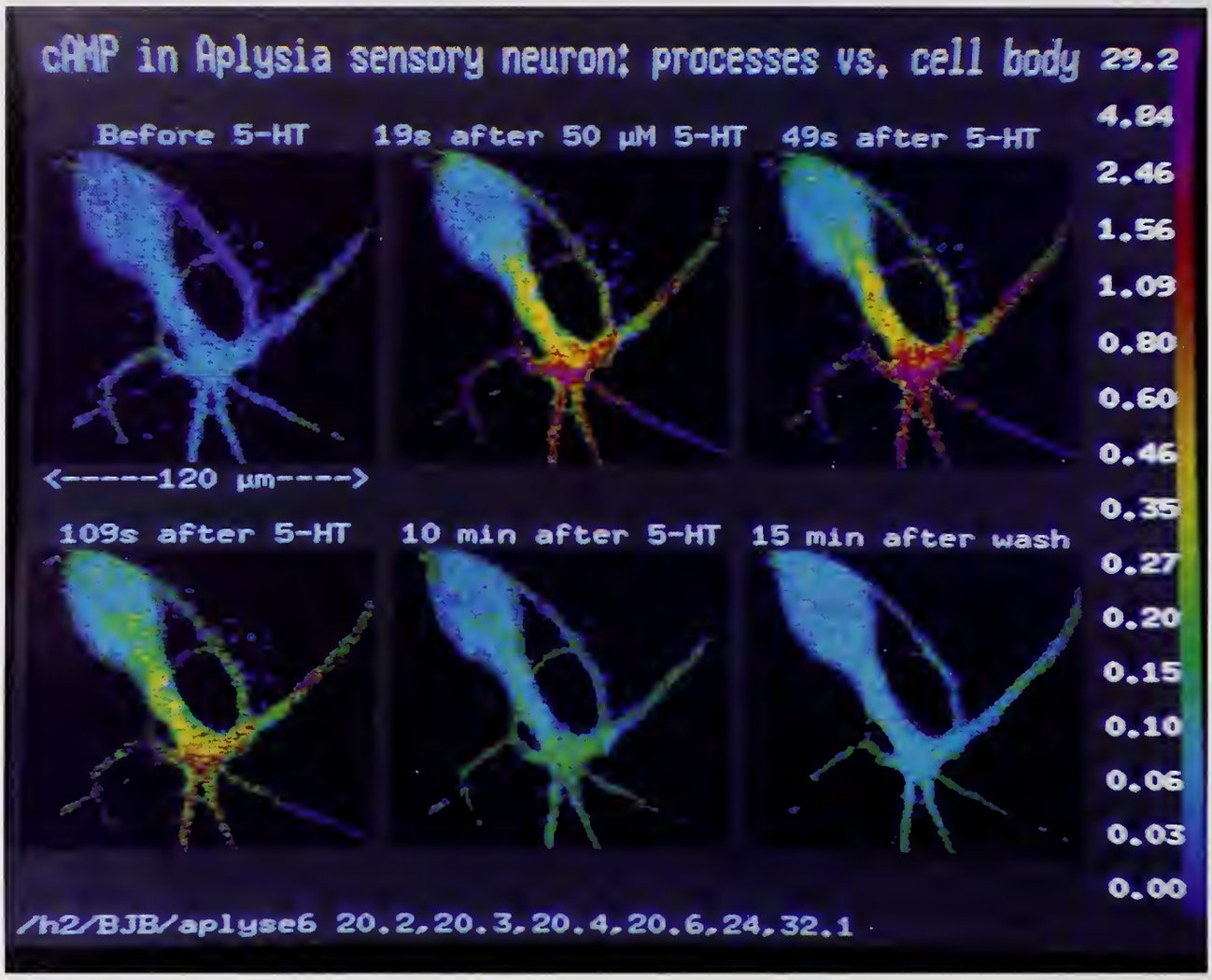
We are also interested in studying how information is stored and retrieved in the brain. The approach that we have taken is to investigate the biochemistry, physiology, and behavior of mice mutant for genes thought to be involved in synaptic plasticity. Using homologous recombination, we have disrupted the α-subunit of the calcium calmodulin kinase II (CaMKII) gene in embryonic stem cells and have used these cells to generate a mouse strain lacking the gene. The α-subunit of CaMKII is neural specific and comprises most of the CaMKII holoenzyme in postnatal hippocampus and forebrain. Peptides that emulate either the calmodulin-binding domain or the inhibitory domain of this kinase seem to block the induction of long-term potentiation (LTP). Mice lacking the α-subunit develop normally and are viable.

Gross neuroanatomical studies did not detect any abnormalities. We studied the electrophysiology of the CA1 fields of the mutant hippocampus in collaboration with the laboratory of Charles Stevens (HHMI, Salk Institute). In normal animals we were able to induce LTP in more than 90 percent (n = 12; 3 animals) of all slices. However, we could stably potentiate less than 10 percent of slices from mutant animals (n = 17; 5 animals). Furthermore, the unpotentiated excitatory postsynaptic potentials of normal and mutant animals were identical, suggesting that synaptic transmission was not impaired in the mutant animals.

We also studied the mutant mice in different versions of the Morris water maze, in collaboration with Jeanne Wehner's laboratory. The results indicate that the mutant mice, but not their normal litter mates, are specifically impaired in tasks that demand the use of configured spatial information. Our analysis of the mice mutant for the α CaMKII demonstrates the involvement of this kinase with LTP and with configural learning.

Past evaluations of the involvement of LTP on learning and memory have mainly depended on the potential role that the NMDA (*N*-methyl-D-aspartate) receptor plays on both phenomena. Unfortunately, NMDA blockers seem to impair both learning and performance, confounding the interpretation of those experiments. However, our results suggest that performance measures unrelated to learning were not responsible for the learning impairment observed in our mutant animals.

A key aspect of our studies has been the apparent specificity of the phenotype: the development and neuroanatomy of the mutant mice are apparently normal, as well as nonpotentiated synaptic transmission. Furthermore, behavioral analysis of the mutant animals argues for a selective impairment on hippocampal-dependent learning. We are continuing to test the Hebbian hypothesis by making mutants for other components of LTP and for genes that might affect the physiology of specific regions involved in learning.



Concentrations of the intracellular messenger cAMP (cyclic adenosine 3′,5′-monophosphate) within a single neuron from the sea snail Aplysia californica. *The cAMP is detected by its ability to affect an enzyme, cAMP-dependent protein kinase. This enzyme was produced by recombinant DNA technology, labeled with fluorescent dyes, and injected into the neuron, which was then grown in tissue culture. The cAMP-sensitive signal from the enzyme was imaged by confocal fluorescence microscopy, a technique that can isolate one plane of focus within a live specimen. The concentrations of cAMP are denoted by a rainbow of colors, with blues and reds representing the lowest and highest levels, respectively.*

The upper left panel shows the neuron before stimulation. The upper middle panel shows the same cell shortly after addition of 5-hydroxytryptamine (5-HT, also known as serotonin) to the medium. Although applied to the entire cell, 5-HT, an important neurotransmitter, elevates cAMP to a much greater extent in the fine dendrites. Over the next 100 seconds (upper right, lower left panels), the cAMP shows some spread to the main cell body, but the dendrites continue to have the highest cAMP levels. The cAMP response begins to decay even while the 5-HT is still present (lower middle panel) and returns to baseline when it is removed (lower right). These images are the first direct visualization of local generation of cAMP in the fine outgrowths and its diffusion toward the cell body and nucleus. It is known to be a key controller of both short- and long-term plasticity in many neurons.

Research of Brian Bacskai, Benny Hochner, Martyn Mahaut-Smith, Stephen Adams, Bong-kiun Kaang, Eric Kandel, and Roger Tsien.

Molecular Engineering Applied to Cell Biology and Neurobiology

Roger Y. Tsien, Ph.D.—Investigator

Dr. Tsien is also Professor of Pharmacology and of Chemistry at the University of California School of Medicine, San Diego. His undergraduate degree was from Harvard College, in chemistry and physics, but it was at the University of Cambridge, England, while obtaining a Ph.D. degree in physiology, that he was "introduced to the potential synergism between organic chemistry and cell biology." After a postdoctoral fellowship at Gonville and Caius College, Cambridge, Dr. Tsien became a faculty member at the University of California, Berkeley. Seven years later his laboratory moved to the University of California, San Diego. His recent honors include the Passano Foundation Young Scientist Award, the Spencer Award in Neurobiology from Columbia University, and the Bowditch Lectureship of the American Physiological Society.

THE overall goal of my laboratory is to gain a better understanding of information processing both inside individual living cells and in networks of neurons. Our preferred approach is through the rational design, synthesis, and use of new molecules to detect and manipulate intracellular biochemical signals, usually by optical means such as fluorescence readout or photochemical release of messenger substances. For example, we have created fluorescent dye molecules that detect calcium ions (Ca^{2+}) with great specificity and sensitivity, so that while the cells are living and performing their normal functions, we can image Ca^{2+} levels inside cells with a spatial resolution of a micron or so and a temporal resolution of a fraction of a second. These dyes have found wide application in cell biology, since a rise in intracellular Ca^{2+} levels is one of the commoner mechanisms by which cell membranes control biochemical events inside the cell, such as muscle contraction, synaptic transmission, glandular secretion, enzyme activation, embryonic fertilization, and growth stimulation.

The detection of intracellular signals such as Ca^{2+} is doubly important. It should help in tracing the complex biochemistries involved in such signaling, and it affords a nondestructive way to watch the activity of many individual cells simultaneously. The latter ability is particularly relevant to understanding how neural networks process information by harnessing many individual but interconnected neurons in parallel. The dominant established techniques for monitoring neural activity either listen intensively to a single neuron at a time or record some smeared-out average of what thousands, millions, or billions of cells are doing. If we can continue to improve the spatial and temporal resolution of present Ca^{2+} imaging, we may succeed in eavesdropping on conversations within small groups of individually identified neurons or in taking snapshots of the instantaneous state of activity of yet larger ensembles. Because imaging is inherently good for following multiple events in parallel, it would be a major help in analyzing the workings of the brain, which is still the most awesome and complex molecular assembly known. We recognize that optical methods, although unsurpassed in their combination of spatial and temporal resolution, are best applied to small regions of thin transparent tissues. For larger volumes of opaque organs, especially in intact organisms, other forms of visualization, such as magnetic resonance imaging, are more appropriate, so we are also seeking to extend our molecular designs to create suitable non-optical indicators.

A recent example of molecular engineering is our development of a fluorescent sensor for cAMP. This important intracellular messenger plays a crucial role in the actions of a great many hormones, in the detection of odors and tastes, and in the mechanisms of learning and memory. In this case we did not design the sensing molecules from scratch but rather modified the natural protein that cells normally use to respond to cAMP. In collaboration with Susan Taylor and her laboratory, Stephen Adams attached fluorescent labels on cAMP-dependent protein kinase in such a way that cAMP not only activates the normal activity of this enzyme but produces an immediate optical signal that we can image microscopically. This labeled protein enables us to visualize cAMP levels, to show that different regions of a single cell can have differing responses to neurotransmitter and drug stimulation, and to see that a subunit of the enzyme can move in and out of the nucleus as the cAMP level rises and falls. While it is in the nucleus, it is ideally placed to modify gene expression.

A particularly dramatic example of the dynamics of cAMP signaling comes from sensory neurons of the marine mollusk *Aplysia californica*. These neurons have been extensively studied by Eric Kandel (HHMI, Columbia University), James Schwartz, and their collaborators as models for both short- and long-term neuronal plasticity.

In collaboration with Dr. Kandel's group, Brian Bacskai and Martyn Mahaut-Smith have injected the labeled protein kinase into the neurons, either in culture or in intact ganglia, and imaged the nucleus, the surrounding cytoplasm of the cell body, and the peripheral outgrowths of the cell. Bath application of the relevant neurotransmitter, 5-hydroxytryptamine, produces rapid increases in cAMP with remarkable spatial gradients—high in the peripheral outgrowths yet only slightly elevated in the central cell body.

Optical sections through the nucleus show that it tends to exclude the holoenzyme (injected into the cytoplasm) as long as cAMP concentrations remain at basal levels. Prolonged elevation of cAMP and dissociation of the holoenzyme causes gradual translocation of the catalytic subunit into the nucleus over tens of minutes. The observed gradient puts high cAMP where it is most needed for short-term plasticity, at the distal processes where the presynaptic terminals would be *in vivo*. Only strong or repeated stimulations would be able to raise the cAMP concentration in the cell body sufficiently to release the catalytic subunit to diffuse into the nucleus, phosphorylate transcription factors, and cause longer-term changes in gene expression. There is some evidence that cAMP changes in the mammalian brain may also be important in comparable forms of plasticity, so we are trying to extend our studies to the appropriate mammalian neurons.

Eventually we hope to extend optical methods to detect macromolecular biochemical signals such as protein phosphorylation or gene transcription. These events currently are assayed by grinding up millions of cells, so that time resolution is limited and differences between individual cells or subregions are impossible to discern. Our experience with imaging ionic messengers and cAMP suggests that cells have considerable individuality and complex behavior patterns. These were missed with destructive population assays, which might be somewhat analogous to studying human psychology on the basis only of anonymous nationwide averages in which the respondents are executed after each poll. We therefore seek continuous, nondestructive readout from single cells. Approaches currently under development (by Julie Matheson and Gregor Zlokarnik, respectively) include microinjection of peptides whose fluorescence is altered by phosphorylation and development of membrane-permeant fluorogenic substrates for reporter enzymes whose nucleotide sequences can be fused to genes or promoter sequences of interest.

A complementary approach is to perturb intracellular signals in a controlled manner to see how the cell or tissue responds. Many important intracellular messengers such as cAMP or inositol phosphates contain one or more phosphate groups that prevent permeability through membranes. A general method for making membrane-permeant analogues has not been available but would be highly useful for artificial stimulation of the putative transduction pathways to see what physiological functions result, especially responses that cannot be assayed in microinjected or permeabilized cells. Recently Carsten Schultz has discovered such a method, esterification of the organophosphate anions with acetoxymethyl groups, which increases the potency of extracellularly applied cAMP analogues by about 100-fold and yields the first membrane-permeant derivatives of inositol phosphates yet reported. This methodology should help us to find new functions for these ubiquitous messengers.

Our projects encompass a wide range of disciplines, including organic synthesis, theoretical and experimental optical spectroscopy and photochemistry, protein chemistry, computerized microscopy and image processing, cell biology, and neurobiology.

The work of Stephen Adams, Martyn Mahaut-Smith, Julie Matheson, Carsten Schultz, and Gregor Zlokarnik in my laboratory was supported by grants from the National Institutes of Health.

Genetic Defects in the Metabolic Pathways Interconnecting the Urea and Tricarboxylic Acid Cycles

David L. Valle, M.D.—Investigator

Dr. Valle is also Professor of Pediatrics, Medicine, Molecular Biology and Genetics, and Biology at the Johns Hopkins University School of Medicine. He received both his undergraduate degree in zoology and his medical degree from Duke University. His internship and residency in pediatrics were completed at the Johns Hopkins Hospital. His postdoctoral research in metabolism was done at NIH.

HUMAN biochemical genetics has been a fruitful area of study since its beginning with the work of Sir Archibald Garrod early in this century. Inherited defects in our body's chemistry or, as Garrod called them, inborn errors of metabolism, are intrinsically interesting and serve as important models for all genetic diseases. My colleagues and I have been involved in the study of several aspects of these disorders, including clinical diagnosis, biochemical characterization, delineation of pathophysiologic mechanisms, development of new therapeutic approaches, and molecular studies of the involved genes.

We have focused on disorders of amino acid metabolism, particularly those involving two fundamentally important areas of metabolism: the urea cycle, which is involved in the conversion of excess nitrogen from a toxic to a nontoxic, readily excreted form; and the tricarboxylic acid cycle, an essential component of energy metabolism. Recently we have extended these interests to include inborn errors in the biogenesis of the peroxisome, a ubiquitous, subcellular organelle that contains about 40 enzymes important in a variety of anabolic and catabolic processes.

One of the amino acid disorders that we are studying intensively is an inborn error of ornithine metabolism known as gyrate atrophy of the choroid and retina (GA). This progressive, blinding chorioretinal degeneration with associated cataract formation is inherited as an autosomal recessive trait. The primary biochemical defect is deficiency of the enzyme ornithine-δ-aminotransferase (OAT), which results in an approximate 10-fold accumulation of ornithine in all bodily fluids.

Despite the systemic nature of the metabolic abnormality in GA, the clinical phenotype is limited to the eye. Thus GA is one of a very few isolated, inherited retinal degenerations for which a primary biochemical defect is known. In an extensive molecular analysis of the OAT genes of 85 probands from GA families around the world, we have detected 34 OAT mutations. Other investigators have added another 20, and together these 54 OAT mutations account for 128 (75 percent) of the possible 170 mutant alleles in our patient population.

This compilation of OAT mutations allows one to determine their consequences on the steady-state levels of OAT mRNA and on the structure and function of OAT protein in the patients' cultured skin fibroblasts or when expressed in a heterologous system, Chinese hamster ovary cells, which lack endogenous OAT mRNA and protein. We find that more than 80 percent of the mutant alleles produce normal amounts of normally sized OAT mRNA. A small fraction (approximately 10 percent) of mutant alleles, all with point mutations that truncate the open reading frame in the penultimate exon or earlier, have markedly reduced levels of OAT mRNA. In contrast to their mRNA phenotype, approximately 80 percent of the OAT mutant alleles, including at least 13 missense mutations, yield little or no detectable OAT antigen. Thus destabilization of the protein is the most common consequence of these mutations. However, two missense alleles, R180T and R154L, inactivate OAT function without reducing OAT antigen. We speculate that the involved residues may play a role in the active site of OAT. Studies are now in progress to produce the large quantities of OAT necessary for x-ray crystallography to determine directly the consequences of these mutations in OAT structure and function. A portion of this work on gyrate atrophy is supported by a grant from the National Institutes of Health.

The OAT-catalyzed reaction is an essential step in the metabolic pathway that interconnects the urea and tricarboxylic acid cycles and, as might be predicted, is subject to complex regulation. In liver, the regulation of OAT expression is coordinated with other urea cycle–related enzymes. We have identified a sequence motif in the 5′-flanking region of OAT that is also present in the promoters of several other urea cycle enzymes and have obtained evidence that this motif is a cis-acting element involved in the regulation of these genes. Surprisingly, localization of OAT expression by *in situ* hybridization and immunohisto-

chemistry reveals an additional complication: OAT expression is limited to one zone of the hepatic lobule, namely a small population of hepatocytes surrounding the central vein, whereas most other urea cycle–related enzymes are in the periportal region. This zonal expression persists even when OAT activity is induced 40-fold by alterations in dietary protein. We hope to identify cis- and trans-acting elements that mediate this aspect of OAT expression. Furthermore, to understand coordinated aspects of OAT regulation better, we are cloning the genes for other enzymes that are metabolically related to OAT. We used complementation in *Saccharomyces cerevisiae* mutants to clone the human cDNA for pyrroline-5-carboxylate reductase, the enzyme that catalyzes the conversion of the product of the OAT reaction to proline. We have now cloned and mapped the structural gene for this enzyme and are beginning a comparison of the promoter regions of the reductase to that of OAT.

Our interest in GA has stimulated us to identify other genes that may be involved in inherited retinal degenerations. We have proceeded along two lines of investigation: 1) cloning genes important for photoreceptor function and 2) using positional cloning to identify candidate genes from a region of the genome known to harbor genes for several retinal degenerations (Xp11.2–Xp11.3). As part of the former strategy, we have cloned the cDNA and cloned and mapped the structural gene for recoverin. Recoverin is a calcium-binding protein whose expression is limited to the photoreceptor. When intraphotoreceptor calcium falls, recoverin stimulates retinal guanylate cyclase, so that photoreceptor cGMP concentrations return to high, dark-adapted levels. We are beginning to examine the possible role of recoverin in a variety of retinal degenerations. In our positional cloning studies of the Xp11.2 region of the human genome, we have assembled yeast artificial chromosome (YAC) contigs covering most of this region and have utilized one of the YACs as a probe to screen a human retinal cDNA library. We have cloned at least five cDNAs, all of which map back to the correct Xp11.2 region. These are being sequenced and will be used as probes in Northern blots of patient samples to test for their possible involvement in these disorders.

We have also begun an investigation of inborn errors of peroxisome biogenesis and function. Zellweger syndrome, a neurodevelopmental disorder fatal in infancy, is the disease paradigm. Cells and tissues from these patients exhibit deficiency of virtually all peroxisomal enzymes and lack normal-appearing peroxisomes. We have cloned the genes for two peroxisomal membrane proteins, the 70-kDa peroxisomal membrane protein (PMP70) and the 35-kDa protein (PMP35). PMP70 is a member of the ATP-binding cassette (ABC) transporter protein family that also includes the mammalian multiple-drug resistance protein (MDR) and the CFTR protein involved in cystic fibrosis. We have cloned the entire PMP70 cDNA, determined its sequence, and used it to clone, map, and characterize the PMP70 gene. In collaboration with Hugo Moser, we are analyzing the possible role of PMP70 in Zellweger syndrome. We have identified three PMP70 mutant alleles, and our results suggest that PMP70 mutations account for one of the Zellweger complementation groups. We also have determined the complete sequence of human PMP35 and identified one mutant allele in another Zellweger complementation group. We now are focusing on expression systems to test directly the functional consequences of these mutations on peroxisomal biogenesis.

Human Molecular Genetics in Two X-linked Diseases

Stephen T. Warren, Ph.D.—Associate Investigator
Dr. Warren is also Associate Professor of Biochemistry and of Pediatrics at Emory University School of Medicine. He received his Ph.D. degree in genetics from Michigan State University. Prior to joining the faculty at Emory, he did postdoctoral research with Richard Davidson at the University of Illinois School of Medicine in Chicago.

MY laboratory is involved in human molecular genetics and is especially interested in the identification of genes responsible for genetic disease. Our efforts currently focus on X-linked disease, particularly the fragile X syndrome and Emery-Dreifuss muscular dystrophy.

Fragile X Syndrome

Fragile X syndrome is the commonest form of inherited mental retardation and one of the most prevalent genetic diseases known, affecting approximately 1 per 1,000 persons worldwide. As the name implies, fragile X syndrome is associated with a fragile chromosomal site, which has been localized to band position Xq27.3. Fragile sites are heritable loci that form cytologically evident gaps within chromosomes under specific biochemical induction. Although such sites are numerous throughout the human genome, the fragile X site is the only one associated with a disease.

Fragile X syndrome is unusual among mammalian genetic disorders in that 20 percent of the males with a fragile X chromosome are not affected while approximately 30 percent of carrier females show some degree of mental impairment. The fragile X mutation is less frequently penetrant (i.e., resulting in mental retardation) among the siblings of these normal carrier males (called transmitting males), and penetrance increases with each generation from a transmitting male until it reaches so-called Mendelian ratios, where half of the male children of a carrier female are affected, typical of an X-linked gene. One exception is that daughters of transmitting males are never affected, but their children may be. This confusing hereditary pattern, unique among genetically studied organisms, has been referred to as the Sherman paradox, in reference to Stephanie Sherman's description of it.

Our work over the past year has not only identified the gene responsible for fragile X syndrome but has also uncovered an unusual mutation whose behavior explains the Sherman paradox. Working with an international group of collaborators, including Thomas Caskey (HHMI, Baylor College of Medicine), David Nelson (Baylor), and Ben Oostra (Erasmus University, the Netherlands), we identified yeast artificial chromosome (YAC) clones, previously developed in my laboratory, that mapped near chromosome breakpoints involving the fragile X site. Using cloned DNA derived from one of these YACs, we identified a cDNA encoded by a gene that the translocation breakpoints had interrupted. This gene, termed *fragile X mental retardation 1* (*FMR-1*), produces a 4.4-kb message expressed at high levels in the brain and testes. Male fragile X patients have macro-orchidism, or enlarged testes, in addition to mental retardation.

Within the *FMR-1* mRNA is an unusual repeat of the trinucleotide CGG. In normal individuals, there are most frequently 29 repeats, though this can vary between 6 and 52. Among transmitting males and most normal carrier females, there are between 52 and 200 CGG repeats. Among mentally retarded patients, the codon repeats up to 1,300 times and is markedly unstable in mitotic cells. Importantly, when the repeat expands beyond 250, it spontaneously methylates the nearby DNA, turning off the *FMR-1* gene. Removal of the gene product by hypermethylation in response to the massive augmentation of CGG repeats is believed to be the mechanism of fragile X syndrome.

Work with our colleagues in Houston revealed an apparent relationship between the number of repeats in a normal carrier female and the probability of having a mentally retarded son. In general, the smaller the abnormal repeat, the lower the risk of expansion to the full fragile X mutation in an offspring. Above a threshold of approximately 200 repeats, the fragile X chromosome when passed down always undergoes expansion to the full mutation. Hence carrier mothers with 200 repeats have a 50 percent risk of having a retarded son, while those with only 70 repeats have a 9 percent risk. This explains the paradox of penetrance in fragile X syndrome: the sequential increase in the CGG repeat with each generation imparts a concomitant increase in risk of having an affected child.

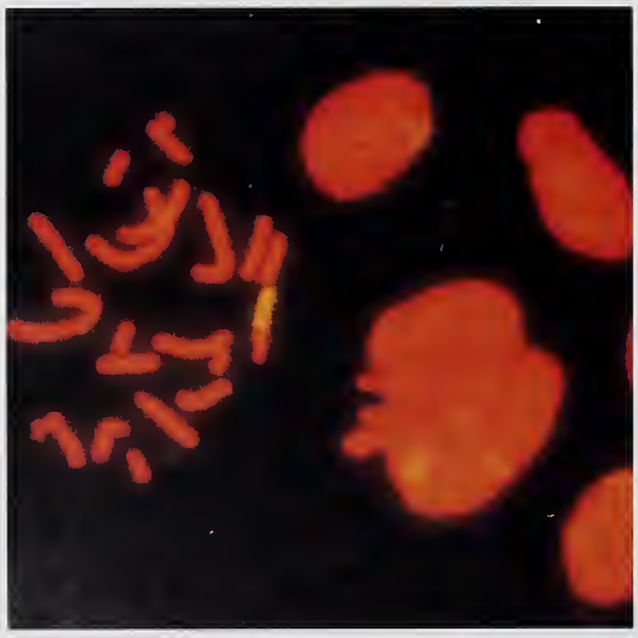
We are focusing on three major efforts related to the *FMR-1* gene. The first is to determine the mechanism of expansion of the CGG repeat by both the direct sequence analysis of the fragile X mutation and the introduction of long synthetic CGG repeats into mammalian cells and, subsequently, into mice. Second, we are attempting to understand the normal function of the *FMR-1* gene product. Toward this goal, we have recently cloned and sequenced the homologous gene from the mouse and are in the process of cloning the yeast and the nematode genes. Finally, we are excited about finding a number of other human genes, distinct from *FMR-1*, that similarly contain long CGG repeats. These genes, whose functions are presently not understood, may share a similar functional utilization of the CGG repeat; may undergo similar mutational changes, perhaps leading to diseases exhibiting unusual genetic patterns; and may represent other, autosomal fragile sites.

Emery-Dreifuss Muscular Dystrophy

Emery-Dreifuss muscular dystrophy (EDMD) is the most frequent X-linked muscular dystrophy following the Duchenne and Becker types. It is a slowly progressive disease that usually leaves affected males ambulatory until middle age. Heart muscle involvement is frequent, sometimes resulting in early sudden death due to heart block. If the disease is identified early, such a death can be prevented by pacemaker implantation.

We have performed genetic linkage studies in two large families and have localized the EDMD gene to the terminal band of the X chromosome (band Xq28) just distal from the fragile X site. We can now place the gene within an approximate 2,000 kb of DNA. Using selective cDNA libraries and DNA of this region cloned into cosmids or YACs, we are mapping muscle-expressed genes within Xq28. This entails use of a somatic cell hybrid mapping panel containing six distinct Xq28 fragments. Any genes that map within the region believed to contain the EDMD gene will be used to search for mutations in patients.

Another outcome of our genetic mapping studies has been the identification of young, asymptomatic males who will eventually suffer from EDMD. Cardiac function of these males is being carefully followed, and pacemaker implantation is performed when warranted prior to full heart block. Thus we have virtually eliminated sudden death due to EDMD in these two families.



In situ *hybridization detecting human DNA within the metaphase chromosomes of a somatic cell hybrid (micro21D), which contains a human-rodent translocation between the centric fragile X chromosome (yellow) and a rodent chromosome arm (red). The hybrid was constructed under conditions favoring rearrangements specific for the fragile X site. Indeed, the human translocation breakpoint in this hybrid is within the CGG repeat of the fragile X mutation. Somatic cell hybrids with translocations of this nature proved instrumental in cloning the fragile X site and its associated gene* (FMR-1).

Research of Stephen Warren.

The MyoD *Gene Family: A Nodal Point During Specification of Muscle Cell Lineage*

Harold M. Weintraub, M.D., Ph.D.—Investigator

Dr. Weintraub is also a Full Member in the Division of Basic Sciences at the Fred Hutchinson Cancer Research Center and Affiliate Professor of Pathology and Zoology at the University of Washington, Seattle. He received his M.D.-Ph.D. degree from the University of Pennsylvania School of Medicine and completed his postdoctoral studies at the Medical Research Council Laboratory of Molecular Biology in Cambridge, England. Before joining the staff at the Hutchinson Center, Dr. Weintraub was in the Department of Biochemical Sciences at Princeton. He is a member of the National Academy of Sciences and the American Academy of Arts and Sciences. Among his many honors are the Eli Lilly Award and the Richard Lounsbery Award from the National Academy of Sciences.

THE *MyoD* gene converts many differentiated cell types into muscle. MyoD is a member of the protein family characterized by the basic helix-loop-helix, a 68–amino acid domain in MyoD that is necessary and sufficient for myogenesis. MyoD binds cooperatively to muscle-specific enhancers and activates transcription. The helix-loop-helix motif is responsible for dimerization, and, depending on its dimerization partner, MyoD activity can be controlled.

MyoD senses and integrates many facets of the cell state. The gene is expressed only in skeletal muscle cells and their precursors; in nonmuscle cells it is repressed by specific genes. MyoD activates its own transcription, perhaps stabilizing commitment to myogenesis. Despite this seemingly overwhelming evidence that *MyoD* is crucial for myogenesis in vertebrates, recent experiments with Michael Krause and with Andrew Fire show that zygotic deletions of *MyoD* in worms result in embryos that retain the capacity to activate muscle cell differentiation.

Muscle-Specific Transcriptional Activation by MyoD

Our laboratory has focused on the mechanism by which MyoD activates transcription. Previous experiments showed that when the 13–amino acid basic region of the ubiquitously expressed basic helix-loop-helix gene *E12* replaces the corresponding basic region of MyoD, the resulting MyoD-E12Basic chimeric protein can bind specifically to muscle-specific enhancers *in vitro* and form dimers with E12, but cannot activate a cotransfected reporter gene or convert 10T½ cells to muscle. Back mutation of this chimeric protein (with the corresponding residues in MyoD) reestablishes activation. A specific alanine is involved in increasing DNA binding, and a specific threonine is required for activation.

A reporter gene containing MyoD-binding sites located downstream from the transcription start site was used to show that MyoD-E12Basic can bind *in vivo* and thereby inhibit expression of the reporter. *In vivo* binding is also supported by the fact that the addition of the "constitutive" VP16 activation domain to MyoD-E12Basic restores full trans-activation potential. The normal MyoD activation domain maps within the amino-terminal 53 residues and can be replaced functionally by the activation domain of VP16.

The activity of the MyoD activation domain is dramatically elevated when deletions are made almost anywhere in the rest of the MyoD molecule, suggesting that the activation domain in MyoD is usually masked. Surprisingly, MyoD-E12Basic can activate transcription in CV1 and B78 cells (but not in 10T½ or 3T3 cells), suggesting that the activation function of the basic domain requires a specific factor present in CV1 and B78 cells. We propose that the masked MyoD activation domain requires, in order to function, the participation of another factor that recognizes the basic region.

By replacing the *MyoD* basic region and the adjacent four-residue junction region with helix 1 into the corresponding region of *E12*, we have recently shown that this small section of *MyoD* is sufficient for myogenesis. Our work suggests that only three residues, A_{114}, T_{115}, and K_{124}, are uniquely critical for "recognition factor" function and subsequent activation of myogenic gene transcription.

Control of MyoD Activity

A variety of transforming agents, including a variety of growth factors, the oncogenes *src*, *ras*, *fos*, *jun*, *fps*, *erbA*, *myc*, and *E1A*, and such chemical agents as butyrate and phorbol esters, inhibit myogenic differentiation. Most of these reagents can inactivate the expressed MyoD protein; in addition, several (such as *ras* and *fos*) also inhibit *MyoD* transcription. Whether this is a secondary

effect due to an inhibition of the autoactivation function of MyoD protein or a more direct inhibition of *MyoD* transcription remains to be determined. Rhabdomyosarcoma cells (derived from tumors of patients who harbor a genetic predisposition to myogenic tumors) differentiate poorly but express MyoD, suggesting that loss of anti-oncogene activity at the rhabdomyosarcoma locus can also impinge on MyoD action. The specific pathway by which each of these oncogenes, anti-oncogenes, and growth factors inhibits myogenesis provides a potential clue to how MyoD might integrate information coming from many aspects of cellular function.

Recently, in collaboration with the laboratory of Inder Verma, we found that the leucine zipper region of the *jun* oncogene actually binds to the helix-loop-helix region of MyoD, both *in vivo* and *in vitro*. Similarly, assays for a recognition factor for MyoD activation show that such a factor, which is missing in rhabdomyosarcoma cell lines, can be provided in trans by fusion with 10T½ cells. Possibly failure to activate myogenesis leads to increased proliferation and then secondary effects that give rise to rhabdomyosarcomas.

Activation of MyoD During Development

We are studying developmental activation of MyoD in mice, worms, and frogs. In both mice and worms, deletional analysis has identified regulatory sequences upstream of the *MyoD* gene that are important for correct developmental activation of MyoD. Current efforts focus on identifying trans-acting elements that integrate with these sequences. In worms, several maternal-effect mutants have been isolated by Jim Priess and his colleagues. These mutant embryos produce excess muscle from the wrong lineage. It is possible that these mutants define elements involved in the segregation of myogenic potential to specific cells during early cleavage stages. In apparent contrast to worms, frogs seem to activate MyoD in all cells of the blastoderm; however, expression is stabilized only in those presumptive mesodermal cells that become induced by vegetal inducing factors such as activin. Frogs also contain maternal *MyoD* mRNA, which, however, seems not to be crucial for subsequent myogenesis, as its destruction with anti-sense DNA results in normal muscle gene activation.

Structural and Functional Studies of the T Cell Antigen Receptor

Arthur Weiss, M.D., Ph.D.—Associate Investigator

Dr. Weiss is also Ephraim P. Engleman Distinguished Professor of Rheumatology and Associate Professor of Medicine and of Microbiology and Immunology at the University of California, San Francisco. He received his undergraduate education at the Johns Hopkins University and was an M.D./Ph.D. student at the University of Chicago, where he studied immunology in the laboratory of Frank Fitch. He did postdoctoral work with Jean-Charles Cerottini and K. Theodore Brunner at the Swiss Institute for Experimental Research, Lausanne. After an internship and residency in internal medicine at UCSF, he became a postdoctoral fellow in rheumatology with John Stobo.

THE immune system has evolved to provide an organism with a flexible and dynamic mechanism to respond specifically to a wide variety of antigens. During the initiation of an immune response, antigen must not only be recognized by antigen-specific lymphocytes, but this recognition event must lead to cellular activation. T and B lymphocytes comprise the antigen-specific components of the cellular immune system. The activation of T lymphocytes is critical to most immune responses, since it permits these cells to exert their potent regulatory or effector activities. During activation, relatively quiescent cells undergo complex changes involving cell differentiation and proliferation.

Following exposure to antigen, activation of T lymphocytes is limited to only those cells expressing antigen-specific receptors. Activation is a consequence of ligand-receptor interactions that occur at the interface of the T cell and an antigen-presenting cell. These interactions initiate intracellular biochemical events within the T cell that culminate in cellular responses. Our goal is to understand how cell surface molecules on the T cell, and in particular the T cell antigen receptor (TCR), initiate T cell activation.

Although it is clear that a number of different cell surface molecules on the T lymphocyte and the antigen-presenting cell may participate in the complex cell-cell interaction that occurs during antigen presentation, the TCR must play a prominent role. Here the familiar lock and key analogy is appropriate. Antigen is the ligand (key) for a particular set of clonally distributed receptors (locks) on T lymphocytes. Antigen often represents a protein fragment that is physically associated with a molecule of the major histocompatibility complex (MHC).

The TCR is an extraordinarily complex structure. It consists of an α/β-chain disulfide-linked heterodimer (Ti) derived from immunoglobulin-like genes that is noncovalently associated with six invariant chains of the CD3 complex and a ζ-chain dimer. CD3 consists of four chains (δ-, γ-, and two ϵ-chains) derived from three closely linked homologous genes located on chromosome 11. The ζ dimer, derived from the products of two homologous genes on chromosome 1, may represent a homodimer ($\zeta\zeta$) or heterodimer consisting of $\zeta\eta$ (η is an alternatively spiced form of ζ) or $\zeta\gamma$ (γ, a homologous protein, is also a component of the IgϵFc receptor on mast cells and basophils). Ti is the ligand-binding subunit of the TCR, since it contains all the information needed to recognize antigen and MHC specificities. CD3 and ζ have been thought to play some role in transducing the ligand occupancy state of Ti across the plasma membrane. Hence the structural basis for the association of Ti, CD3, and ζ is of interest.

Previous studies from our laboratory have demonstrated that coexpression of all of the chains of the oligomeric TCR—Ti, CD3, and ζ—on the plasma membrane is obligatory for efficient TCR expression. Studies from this laboratory demonstrated that the structural and functional basis for the interaction between Ti, CD3, and ζ is contained within regions including the transmembrane domains of these proteins. Further mutational studies are in progress to understand more precisely how Ti, CD3, and ζ interact functionally within these domains.

Since the transmembrane regions are responsible for the Ti-CD3-ζ association, we have taken advantage of this information to separate regions or domains of CD3 and ζ from Ti. This has been accomplished by constructing chimeric molecules between other cell surface molecules linked to the cytoplasmic domain of TCR chains. A chimeric molecule consisting of the CD8 extracellular and transmembrane domains fused to the ζ cytoplasmic domain acquired the signal-transducing capacity of the entire TCR. This finding demonstrates that the cytoplasmic domain of CD3 ζ can link the TCR to intracellular signaling machinery.

These studies are being extended to define the region of ζ that interacts with such intracellular molecules and to identify these molecules. Recently we identified a 70-kDa tyrosine phospho-

protein that associates with the ζ-chain; we are in the process of characterizing this protein. We would also like to understand the function of the associated CD3 complex within the complex oligomeric TCR. Moreover the ability to create such functional chimeric receptors may permit the creation of novel antiviral or antitumor TCRs that may be of value in gene therapy.

Stimulation of the TCR initiates cellular activation by inducing a transmembrane signal that is manifested as the formation of intracellular biochemical mediators called second messengers, which can initiate or influence cellular response pathways. Recent studies demonstrate that the TCR activates an uncharacterized protein-tyrosine kinase (PTK) as the initial event leading to cellular activation. This PTK appears to be associated with the TCR ζ-chain. The activation of this PTK results in the tyrosine phosphorylation of many cellular proteins, one of which is phospholipase C-γ1 (PLC-γ1). The phosphorylation of PLC-γ1 activates this enzyme to hydrolyze a rare membrane lipid, PIP_2 (phosphatidylinositol 4,5-bisphosphate). This yields two potent intracellular second messengers (inositol 1,4,5-trisphosphate and diacylglycerol) that regulate the mobilization of intracellular calcium and activation of the enzyme protein kinase C, respectively. These latter two events are physiologically important to subsequent cellular responses.

The mechanism by which the TCR couples to intracellular signaling pathways is largely undefined, as are many of the components of the signaling pathways themselves. To define and characterize the molecular basis by which the TCR regulates these pathways, we are using a somatic cell genetic approach. In work supported by a grant from the National Institutes of Health, we have isolated a number of mutants derived from T cell leukemic lines that are defective in TCR-mediated activation of the inositol phospholipid pathway. Unlike the parental cells, none of these mutants produce lymphokines in response to TCR stimulation. These mutants define four distinct gene products other than the TCR chains that are required for the functional activation of the inositol phospholipid pathway.

In one mutant, the activation of the tyrosine kinase pathway is still intact; however, some of the substrates of the tyrosine kinase pathway are not phosphorylated, including PLC-γ1. The phenotype of this mutant suggests that there may be a coupling protein or second PTK that is deficient.

The defect in another of these mutants can be attributed to the absence of a cell surface protein, CD45, with tyrosine phosphatase activity. The absence of CD45 prevents the TCR from activating the tyrosine kinase or phosphatidylinositol inositol pathway. At least one target of the CD45 tyrosine phosphatase is a regulatory phosphotyrosine residue in the PTK lck. Thus a protein-tyrosine phosphatase regulates the activity of a PTK. This suggests a complex autoregulatory system that we are intensively studying. In the remaining two mutants, biochemical studies have complemented our genetic approach. In neither of these mutants is the PTK pathway activated. Preliminary studies suggest that one of these mutants is deficient in a previously identified PTK. Thus these mutants are proving to be valuable tools with which to dissect the complexities of the signal transduction pathways and their relationships to cellular responses.

T cell activation is a complex process that is regulated by cell surface molecules. Investigation of the molecules and events involved in the activation of T cells should lead to a more complete understanding of T cell biology and a more rational approach to the manipulation of the immune system. Moreover, through the study of the activation of T cells, it is likely that insight into other biological systems involving cell proliferation and differentiation will emerge.

Following the Life History of Lymphocytes

Irving L. Weissman, M.D.—Investigator

Dr. Weissman is also the Karel and Avice Beekhuis Professor of Cancer Biology and Professor of Pathology, Developmental Biology, and (by courtesy) Biology at Stanford University School of Medicine. He directs the Program for Molecular and Genetic Medicine and the Immunology Program. He received his M.D. degree from Stanford and remained to do postdoctoral studies in the Department of Radiology. He also studied at Oxford with Jim Gowans in 1964 and returned in 1975 for part of a sabbatical year, which he then completed with Melvin Cohn at the Salk Institute. Dr. Weissman is a member of the National Academy of Sciences and the American Academy of Arts and Sciences.

LIKE all other blood cells, lymphocytes—the principal players in immune recognition of self from nonself—are derived ultimately from stem cells in the bone marrow. It is both biologically and clinically important to delineate the decisions these bone marrow precursors make as they pass through microenvironments that define the type of lymphocyte (or other blood cell) they shall become. We have focused on identifying the earliest cells in mouse and human bone marrow that have multipotent capacity, the so-called hematopoietic (blood-forming) stem cells.

Several years ago we were able to isolate the hematopoietic stem cell of the mouse. This year we showed that no other cell type in the bone marrow has stem cell activity or potential.* We have also demonstrated its full developmental potential by transferring a single stem cell from one mouse strain mixed with 100 stem cells from another strain into lethally irradiated mice of the second strain. Progeny from the single marked stem cell regularly gave rise to over 100 million blood cells of all types, including hundreds to thousands of stem cells.*

These thousands of stem cells, derived from the initially injected single cell, could be retrieved and transferred to a second generation of irradiated animals, all of whom were fully reconstituted. Thus this stem cell has a remarkable profile of activities, including that of massive self-renewal.

In the past year we also found that stem cells in mouse fetuses have the capability of giving rise to a broader variety of T cells than do stem cells in the adult bone marrow.* The additional types of T cells derived from fetal stem cells are those cells that move from the fetal thymus to the skin and other epithelial coverings of the body, presumably to act as sentinels to protect against incoming infectious microorganisms.

Most remarkably, in preparation for this added capacity in fetal life, stem cells apparently start a T cell developmental "clock" on one T cell receptor chromosome. We propose that after several stem cell divisions, the clock moves past that part of the chromosome that will be expressed as receptors for antigens expressed on epithelial T cells, and concentrates only on other T cell receptors to fight off infection in other sites of the body. We believe that the genetic events that set developmental clocks and then shut them down at the level of hematopoietic stem cells lie at the heart of understanding the determination of choices that cells can make in general, and hope to develop in the next few years new methods to investigate the regulators that set the clock and how they do so in the developmental history of the mouse embryo.

Several years ago our laboratory developed a mouse model of human organ function, wherein human hematopoietic and lymphoid organs, such as fetal liver, thymus, and bone marrow, could be implanted in the immunodeficient SCID mouse.* We found that the human lymphoid microenvironments implanted in the SCID mouse could provide the right soil for human T cell lymphoma growth, while the same primary tumors from patients will not grow in the SCID mouse in any other microenvironment.* This presents the opportunity to study the earliest stages of growth and malignant progression of human cancers, lymphomas, and leukemias, if the general principle holds that their early growth is dependent on the organ in which they find themselves. This implies that there might exist factors within human organs that are responsible for the early neoplastic growth of cancer cells, some of which may be tissue specific.

In the past year, we also identified the genes that encode a Peyer's patch homing receptor—the molecule that is involved in the traffic of lymphocytes from the bloodstream to intestinal lymphoid organs such as Peyer's patches, appendix, and mesenteric lymph nodes. The molecule is a member of the integrin family of adhesion proteins and uses the combination $\alpha_4\beta_7$. Lymphomas that express $\alpha_4\beta_7$ were found to metastasize to

* This work is supported by a grant from the National Institutes of Health.

Peyer's patches as well as to bind to the blood vessels in the traffic zones of Peyer's patches, whether the $\alpha_4\beta_7$ molecules were found to be expressed on these lymphomas or were transfected to them by our cDNA clones. Antibodies to either α_4 or β_7 could block this selective adhesive event.

We have also developed a new mouse strain by transgenic technology, deleting all cells that express the activated killer cell gene granzyme A (first identified and cloned in this laboratory) by attaching to that gene a "suicide" gene obtained from Richard Palmiter (HHMI, University of Washington). When killer cells are activated in this mouse strain they begin to express the diphtheria toxin A chain suicide protein, and die. Unexpectedly, we have uncovered in this transgenic mouse strain a profound effect on the life span of all CD8 T cells, and not just killer T cells and natural killer cells. In the next year, we plan to utilize these mice to delineate the role of killer cells in normal and pathological immune reactions *in vivo* and to define the mechanism by which non–killer CD8 T cells are deleted after their emigration from the thymus.

Function and Regulation of the Cystic Fibrosis Transmembrane Conductance Regulator

Michael J. Welsh, M.D.—Investigator

Dr. Welsh is also Professor of Internal Medicine and of Physiology and Biophysics at the University of Iowa College of Medicine, Iowa City. He earned his M.D. degree from the University of Iowa. He completed his residency at the University of Iowa College of Medicine; held clinical and research fellowships in pulmonary diseases and cardiovascular research at the University of California, San Francisco; and did postgraduate research in physiology and cell biology at the University of Texas, Houston. He then returned to the University of Iowa as a faculty member.

CYSTIC fibrosis (CF) is a common lethal genetic disease involving defective electrolyte transport by several epithelia. In normal epithelia of the airways, the intracellular second messenger cAMP regulates chloride (Cl^-) channels in the apical membrane. These channels provide both a pathway through which Cl^- flows and a key point for regulation of its movement. When cAMP increases, the channels open and Cl^- flows from the cell into the airway lumen, drawing water with it. Secretion of salt and water is important in generating the respiratory tract fluid, a critical component of the mucocilliary defense mechanism. In CF airway epithelia, cAMP fails to open the Cl^- channels. As a result secretion is defective and the respiratory tract fluid is abnormal. This defect is believed to be the major cause of morbidity and mortality in CF lung disease.

CF is caused by mutations in the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR). When we expressed the normal CFTR gene in CF airway epithelial cells, the channel defect was corrected. That result indicated that CFTR was somehow intimately associated with Cl^- channels, but did not reveal its function. Toward further understanding, we expressed CFTR in a number of nonepithelial mammalian cells lacking both endogenous CFTR and cAMP-activated Cl^- channels. In every case in which CFTR was expressed, the cell generated a unique Cl^- channel that was activated by cAMP. Such channels were not observed in cells lacking the CFTR gene.

The cAMP-regulated Cl^- channels displayed regulatory and biophysical properties that were identical to those observed in cells expressing endogenous CFTR, as well as those in the apical membrane of airway epithelia. We stress apical membrane because that is where the CF defect is observed. These results suggested that CFTR itself might be a Cl^- channel. This conclusion was quite controversial because CFTR did not resemble any channels previously described. It seemed instead to resemble a family of proteins that include membrane pumps.

To test the hypothesis that CFTR is a Cl^- channel, we changed specific amino acids within the CFTR sequence. When an ion crosses a cell membrane through a channel, it must interact with the amino acids in that channel. In changing some of the positively charged amino acids to negatively charged ones, we changed the anion selectivity sequence. The channels normally favor Cl^- over iodide (I^-), but after two of the amino acids were mutated in CFTR, the channel favored I^- over Cl^-. The ability to change the properties of the conduction mechanism by altering specific amino acids provided the most compelling evidence that CFTR is itself a cAMP-regulated Cl^- channel.

The evidence that CFTR forms a Cl^- channel and that the electrolyte transport defect in CF is in the apical membrane suggested that CFTR would be located in the apical membrane of secretory epithelia. To test that hypothesis, we developed antibodies to CFTR and localized it with immunofluorescence confocal microscopy. In several lines of intestinal epithelial cells that secrete Cl^-, we found that CFTR was located in the apical membrane. That result indicates that CFTR is in a position where it can directly mediate Cl^- movement across the membrane.

An increase in the cellular concentration of cAMP opens the CFTR Cl^- channel. Several of our studies have shown that CFTR is phosphorylated and thus regulated by a cAMP-dependent protein kinase (PKA). (Kinases regulate cell proteins by attaching a phosphate group.) When we used cell-free patches of membrane containing CFTR, we found that addition of the catalytic subunit of PKA opened the CFTR Cl^- channel.

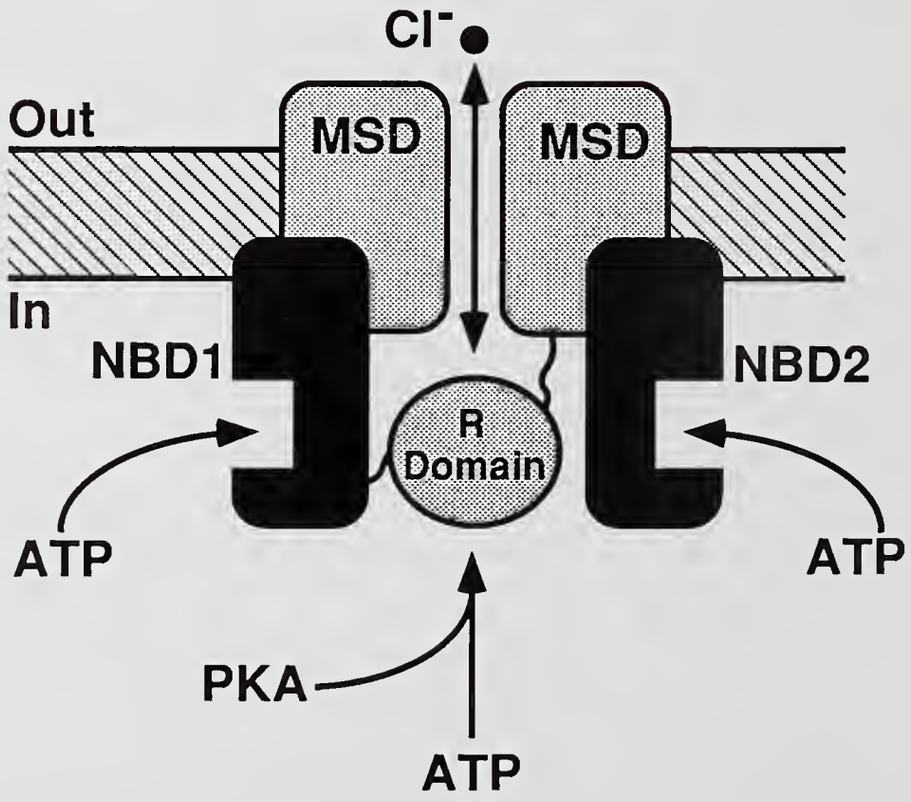
Moreover, *in vitro* biochemical studies disclosed that PKA phosphorylates CFTR on seven residues. *In vivo* studies further showed that PKA phosphorylates four serine residues located within a portion of the protein called the R domain. Evidence that these reactions are important for opening the channel came from the observation that PKA failed to open the channel when the four serines had been changed to alanines. Furthermore, when the R domain was deleted from CFTR, the channel no longer required PKA to open: it was open even without phosphorylation.

The CFTR Cl^- channel also contains stretches

of amino acids similar to those found in a number of proteins that utilize ATP. To determine if the CFTR Cl^- channel utilizes ATP, we first phosphorylated it with PKA and then examined the effect of ATP on cell-free membrane patches containing CFTR. We found that ATP is required for the channel to open. The channel appears to require the hydrolysis of ATP; nonhydrolyzable analogues would not open it. This observation provides the first example of an ion channel regulated by ATP hydrolysis. We were surprised because an ion channel, by its nature, is a passive structure. The results suggest that the energy from ATP hydrolysis may be required to open the channel and that, once open, Cl^- flows through passively.

These studies have begun to define the function of CFTR, and in so doing, they begin to explain why the apical membrane of CF epithelia are impermeable to Cl^-. They also suggest that correction of the underlying defect would be a feasible therapeutic strategy.

In addition to the studies of the CFTR Cl^- channel, the laboratory is investigating other channels in epithelial cells. Other Cl^- channels are not as well explored as the CFTR channel; but if their regulation were understood, they might possibly be utilized to bypass the CFTR Cl^- channel defect. Potassium channels, too, are important, because they generate the intracellular voltage that drives Cl^- out of the cell across the apical membrane into the airway lumen. Finally, the laboratory has a major focus on the regulation of intracellular calcium. The calcium ion is known to regulate some Cl^- channels in the airway epithelial cells, and again, a better understanding of this process might open an approach by which the CF defect could be circumvented.



Model showing the proposed domains of the cystic fibrosis transmembrane conductance regulator (CFTR). Membrane is cross-hatched area. MSD refers to membrane-spanning domain, NBD to nucleotide-binding domain, and PKA to cAMP-dependent protein kinase. Mutation of the CFTR gene is responsible for the defective function of chloride channels in CF patients.

Adapted from Anderson, M.P., Berger, H.A., Rich, D.P., Gregory, R.J., Smith, A.E., and Welsh, M.J. 1991. Cell *67:775–784. Copyright © 1991 by Cell Press.*

Identification of the Gene Responsible for Adenomatous Polyposis

Raymond L. White, Ph.D.—Investigator

Dr. White is also Professor of Human Genetics and Adjunct Professor of Cellular, Viral, and Molecular Biology at the University of Utah School of Medicine. He received a B.S. degree in microbiology from the University of Oregon, Eugene, and a Ph.D. degree in microbiology from the Massachusetts Institute of Technology. He did postdoctoral research with David Hogness at Stanford University. Dr. White has held various academic appointments at the University of Massachusetts, Worcester, and the University of Utah. He was recently named the Thomas D. Dee II Professor of Human Genetics at the University of Utah. Dr. White is a corecipient of the National Health Council's National Medical Research Award and was recently elected to the National Academy of Sciences.

FORMATION of a polyp in colonic epithelium is an early event in the development of colon cancer. In some families a dominantly inherited mutation causes familial adenomatous polyposis coli (APC), a condition characterized by large numbers of polyps and a consequently high risk for colon cancer among those who inherit the mutant allele. Because colon carcinoma is the leading cause of cancer death in the United States, identification of the gene that causes polyposis in APC families has been the goal of investigators seeking to understand cellular mechanisms that can lead to malignancy. This goal was achieved when the gene *APC* was isolated in 1991.

Genetic linkage studies, combined with evidence from microscopic analysis of altered chromosomes, had localized the polyposis-causing mutation in APC families to a specific region on the long arm of chromosome 5. Molecular techniques, including physical mapping of large DNA fragments, were brought into play to define the region more clearly, because numerous genes besides *APC* were likely to reside in that portion of the chromosome.

With molecular probes derived from the large DNA fragments, small deletions were found in the DNA of two unrelated patients with APC. These deletions proved to be the key to isolation of *APC*, for they defined a very small region in which to search. A nearby gene, *MCC*, which is often mutated in sporadic colonic tumors, had been a candidate; but *MCC* was not interrupted by either of the small deletions, and the search continued.

Three other genes were found that did fall within the 100-kilobase region deleted in both patients, and each of those was investigated for the presence of very small ("point") mutations in the DNA of other persons with APC. One of these genes was subsequently identified as *APC* on the basis of point mutations in several patients.

Although each of these APC-specific genetic alterations had occurred in a different coding region of the gene, all were of a kind that could be expected to prevent the generation of a normal protein product. In one patient the mutation was new in his family. His parents' chromosomes carried normal copies of *APC*, but two of his children had inherited his mutant allele.

APC, like *NF1* (neurofibromatosis type 1), another tumor-associated gene cloned recently, appears to fall into the growing category of tumor-suppressor genes. There were no clues, however, to its physiological function, as had come to light for *NF1* with the discovery of similarities in amino acid sequence between its predicted product and proteins known to participate in transduction of growth signals within the cell. The only structural motif that could be identified was the presence of "heptad repeats," or series of seven amino acids in tandem arrays. Heptad repeats often suggest that a protein can interact with similar proteins to form coiled helical structures.

The putative product of *MCC* also contains heptad repeats. It is intriguing to speculate upon the possibility of interaction between this protein and the *APC* product, in view of the fact that alterations in both genes are known to play important roles in the pathway to colon cancer.

Some families exhibit an unusually high incidence of colon cancer without showing a pattern of classical polyposis; that is, affected members may develop very few polyps. Inherited mutations in the *APC* gene are being sought in these families, and colonic tumors in the general population are being tested for mutant *APC* alleles. Characterization of a variety of *APC* mutations may help to explain differences in clinical profiles among people with a genetic predisposition to colon cancer. It may also lead to new information about the role of the gene in normal cellular processes.

For many families with APC, discovery of the gene has immediate value for presymptomatic, even prenatal, diagnosis of carrier status. When a

particular mutation has been characterized in a family, members who carry the disease allele can be identified directly. Families for whom the mutant allele is as yet unidentified may still benefit from analysis of genetic linkage between the putative disease allele and a highly polymorphic DNA marker system that has been detected within one of the other genes lying in the *APC* deletion region. Early diagnosis at the molecular level will free unaffected members of these families from some anxiety and spare them the discomfort and expense of frequent colonoscopic examinations.

Mechanisms of the Biological Activities of Membrane Glycoproteins

Don C. Wiley, Ph.D.—Investigator

Dr. Wiley is also Professor of Biochemistry and Biophysics at Harvard University and Research Associate in Medicine at the Laboratory of Molecular Medicine at the Children's Hospital, Boston. He received his Ph.D. degree in biophysics from Harvard University. He then joined the faculty at Harvard and served as Assistant and Associate Professor of Biochemistry and Molecular Biology before attaining his present position. Dr. Wiley is Chairman of the Department of Biochemistry and Molecular Biology. He is a member of the National Academy of Sciences and a Fellow of the American Academy of Arts and Sciences. Among his honors is the Louisa Gross Horwitz Prize from Columbia University.

T cell recognition occurs when cell surface histocompatibility glycoproteins present antigens, processed to small peptides, to an antibody-like molecule on the T cell receptor. An individual organism has only a small number of different histocompatibility molecules (probably less than a dozen), so that each histocompatibility glycoprotein must be able to "present" many, possibly thousands, of different antigenic peptides to thousands or more distinct T cell receptors throughout the immunological life of the individual.

In the past year we have been able to visualize the conformation of peptides bound to a histocompatibility glycoprotein, HLA-B27. The extended conformation of the peptide appears to be specified by the HLA-binding site, so that the two ends of the peptide are bound to specialized regions at the two ends of the binding groove. A few of the side chains of a 9-mer peptide interact with pockets in the surface of the HLA molecule, whose size and chemical composition vary from allele to allele in the population. We also eluted, sequenced, and identified 11 self-peptides from HLA-B27. All 11 had arginine, a positively charged amino acid, at peptide position 2, which correlates with the x-ray crystallographic finding that position 2 fits into a deep pocket with a negatively charged glutamic acid at the bottom. (That residue is polymorphic and changes the specificity of that pocket in other alleles.)

We are also now able to reconstitute class I molecules from polypeptide chains produced in *Escherichia coli* with single peptides and have crystallized HLA-A2 with a series of peptide antigens from influenza virus and HIV-1 (human immunodeficiency virus type 1). One complex diffracts beyond 1.5-Å resolution when the crystal is frozen at −160°C. (All of the crystals in our laboratory are now frozen to this temperature to preserve crystallographic order and eliminate radiation damage.)

In our studies of class II histocompatibility antigens (a collaboration with Joan Gorga and Jack Strominger), we now have three crystals under study: human DR1, human DR1 plus a superantigen, and DR1 expressed in insect cells and complexed with a single influenza virus peptide. The complex of a single peptide with a class II molecule was generated by expressing a soluble class II molecule in cells from insects, which lack an immune system. Empty DR1 molecules were produced that rapidly and stoichiometrically bound peptide. The empty molecules were stabilized against aggregation and sodium dodecyl sulfate–induced denaturation by addition of peptide, arguing that peptide binding is accompanied by a conformational change.

Our laboratory is also studying how influenza virus infects cells. About 10 years ago we determined the three-dimensional structure of the influenza virus hemagglutinin (HA), the viral glycoprotein responsible for binding the virus to cells and for fusing the viral membrane to a cellular membrane to effect infectious entry. Recently we determined the structure of a series of complexes between the HA and derivatives of sialic acid, the cellular receptor for influenza virus. We have synthesized a number of these new ligands and determined the crystal structure of the complexes to confirm an atomic model for virus-cell binding that we proposed. In the process a second binding site has been located on the HA at an interface between domains of the molecule, which, although probably not physiological, may offer opportunities for the design of a ligand to stabilize the interface against the conformational change required for the HA's membrane fusion activity.

A number of other crystallographic and biochemical studies are under way on influenza C virus, on a low-pH fusion-active conformation of the influenza HA, on trypanosome surface antigens, and on the glycoprotein of HIV-1 in complex with its cellular receptor, CD4.

Studies of Blood Cell Formation

David A. Williams, M.D.—Assistant Investigator

Dr. Williams is also Associate Professor and Kipp Investigator of Pediatrics at Indiana University School of Medicine, Indianapolis. It was here that he received his medical degree, after graduating in biology from Indiana State University. His postdoctoral training includes a pediatric residency at Children's Hospital Medical Center, Cincinnati; research fellowships with Richard Mulligan at the Massachusetts Institute of Technology Center for Cancer Research and the Whitehead Institute for Biomedical Research; and clinical fellowships in pediatric hematology/oncology at Harvard Medical School, the Children's Hospital, and Dana-Farber Cancer Institute, Boston. He was Assistant Professor of Pediatrics at Harvard Medical School before assuming his present position.

THE goal of work in our laboratory is to understand the relationship between supporting cells in the bone marrow cavity and the development of blood cells. The production of blood cells from primitive "stem cells" in the marrow is dependent on interactions between the latter cells and certain supporting cells. Understanding the process of blood cell formation is important for treatments that utilize bone marrow transplantation, including gene transfer methods to correct genetic diseases affecting the blood-forming system.

Studies of the Bone Marrow Environment

The production of enormous numbers of blood cells in the bone marrow environment and their delivery to the blood circulation constitute a highly regulated and complex process. Either excessive or insufficient production of the primitive stem cells or the more mature daughter cells is associated with human diseases. Little is known, however, about how this process is regulated. Our approach to the problem has been to simplify interactions between stem cells and supporting cells by immortalizing single supporting cells from the bone marrow and studying these interactions in detail.

In collaboration with Vikrum Patel at Northwestern University, we have defined a protein and its receptor that appear to be important in anchoring stem cells to supporting cells. This interaction can be blocked in mice and *in vitro* using antibodies to the receptor. The stem cells that adhere to supporting cells (stromal cells) using this receptor appear to have a high capacity to form new stem and daughter cells. The adherence of stem cells to supporting cells is important in the survival of stem cells both *in vitro* and in the bone marrow cavity. A better understanding of this interaction may lead to new methods for growing blood cells and expanding the number of stem cells in the laboratory.

An important protein in the growth of blood cells was recently identified in our laboratory as "Steel factor." This protein is defective in a mouse mutant (Steel mouse), which has a form of aplastic anemia, as well as infertility and skin abnormalities. The protein is normally made in two forms that differ with respect to how it is presented to blood stem cells in the bone marrow environment. We are investigating the role of these two different types of Steel factor presentation.

Deniz Toksoz, a former HHMI associate, has recently shown that presentation of Steel factor in a localized fashion leads to survival of immature human blood cells (termed progenitor cells) longer in culture than when the presentation is nonlocalized. This observation, in combination with the identification of proteins involved in blood cell adhesion (described above), may have important implications in developing a system for growing bone marrow cells *in vitro*. Manas Majumdar, an HHMI associate, is currently studying the role of this protein on blood formation during development of the mouse.

Studies on the role of the supporting cells in the bone marrow have also led us to the identification and production (by recombinant methods) of another protein involved in blood formation, termed IL-11 (interleukin-11). In recent work utilizing IL-11, Xunxiang Du has demonstrated that it has the capacity to help bone marrow stem cells reestablish blood formation after intravenous injection (i.e., bone marrow transplantation) in mice. This protein affects the growth of several different types of blood cells, including platelets and neutrophils. We are investigating the use of this growth factor to prevent some of the severe complications of chemotherapy, such as infection and bleeding. This work is also supported by grants from the National Institutes of Health.

Effects of Growth-regulating Proteins During Blood Development in the Early Embryo

In the developing mammalian embryo, blood

formation begins in the yolk sac. Little is known about the interaction between supporting cells in this environment and the earliest primitive stem cells, a very few of which are thought to give rise to all blood cells in the adult. Our laboratory has derived cell lines from each part of the yolk sac, and Merv Yoder, a visiting scientist, is characterizing the interactions of supporting cells from the yolk sac with blood stem cells derived from mouse adult bone marrow, yolk sac, and the embryonic liver.

In addition, Pamela Kooh, an HHMI associate, is examining the effects of expression of various growth proteins during fetal development. She is utilizing embryonic stem cell technology to examine mouse fetal blood development *in vivo* and *in vitro*.

Gene Transfer and Somatic Gene Therapy

A long-standing focus of research in our laboratory has been the use of viral vehicles, or vectors, to introduce new genetic material into bone marrow stem cells. These manipulated cells can then be introduced into a recipient by bone marrow transplantation, so that the daughter blood cells contain the new gene. In the future these methods, called somatic gene therapy, may provide a means of curing severe genetic diseases.

Our work has utilized information we have gained about blood formation to improve the delivery of genes into mouse stem cells. Barry Luskey, an HHMI associate in our laboratory, has shown that use of the Steel factor ensures that 100 percent of mice transplanted with manipulated bone marrow can be made to express human protein. We hope to apply this knowledge clinically. Children with adenosine deaminase (ADA) deficiency, a genetic disease involving bone marrow–derived cells, exhibit severe combined immunodeficiency (SCID). With a view to treating this fatal condition, we are collaborating with David Bodine and Arthur Nienhuis at the National Institutes of Health in testing the approaches outlined above in monkeys, but so far without success.

Tom Moritz, a postdoctoral fellow in our laboratory, is collaborating with Ronald Hoffman at Indiana University in investigating the transfer of the ADA gene in human bone marrow cells that have been highly enriched for stem cells. At this point it is unclear whether the use of such stem cell enrichment will improve the efficiency of gene transfer into blood stem cells. Gene transfer efficiency is a critical issue in the successful extension to humans of gene transfer methods developed in mice by this and other laboratories.

Growth Factor–stimulated Cell Proliferation

Lewis T. Williams, M.D., Ph.D.—Investigator

Dr. Williams is also Professor of Medicine at the University of California, San Francisco. He received his undergraduate degree from Rice University and his M.D. and Ph.D. degrees from Duke University, where he studied with Robert Lefkowitz. He then completed a clinical residency in internal medicine and specialty training in cardiology at Massachusetts General Hospital, Boston. Before joining the faculty at UCSF, he was Assistant Professor of Medicine at Harvard Medical School. Among his honors is the Outstanding Young Investigator Award of the American Federation for Clinical Research.

POLYPEPTIDE growth factors regulate the proliferation and migration of cells in the developing tissues of embryonic animals. The actions of these factors appear to be recapitulated in adults when damaged or senescent tissues are repaired. Our research group is investigating the action of platelet-derived growth factor (PDGF). This potent growth factor for fibroblast and smooth muscle cells is found in platelets and is released at sites of tissue injury. PDGF is also produced by other tissues, including endothelial cells that line the inner surfaces of blood vessels. In this context PDGF is likely to play a major role in stimulating the proliferation of smooth muscle cells that constitute atherosclerotic plaques. Its role in vascular proliferation appears to be especially prominent in the recurrent blockage of coronary arteries that occurs after clinical interventions such as angioplasty or atherectomy, which are undertaken in an attempt to restore blood flow through vessels narrowed by atherosclerosis. PDGF also plays a role in the growth of some tumors. At least one monkey sarcoma is caused by the aberrant production of PDGF, which stimulates the tumor cells to grow in an uncontrolled fashion.

Like other growth factors, PDGF acts on cells by first binding to specific receptor sites located on the cell surface. This interaction of PDGF with its receptor is transmitted as a signal across the cell membrane and triggers a series of complex reactions inside the cell that culminate in DNA synthesis and cell division. To study the mechanism of signal transmission by the PDGF receptor, we purified the receptor from mouse cells, cloned the gene that encodes the mouse receptor protein, expressed this receptor in cells that normally lack PDGF receptors, and demonstrated that this expressed receptor mimics the actions of native PDGF receptors and mediates all of the known responses to PDGF.

The PDGF receptor is anchored at the surface of the cell and is oriented so that approximately half of the receptor, the PDGF-binding domain, is located outside the cell, and the other half of the receptor is located inside the cell. The receptor appears to consist of several distant regions, termed domains, that have distinct functions. Using the cloned gene for the receptor, we can produce individual domains of the receptor protein and study the functions of these domains. For example, we have produced receptor fragments that contain the PDGF-binding domain but lack the other portions of the molecule. To localize more precisely the portion of the receptor that is essential for binding PDGF, we are now preparing even smaller versions of the binding domain by deleting portions of the molecule. When a minimal domain for binding is defined, we will study the three-dimensional structure of this simplified molecule and use this information to design agents that can block the binding of PDGF to its receptor. These agents should function as blockers of the actions of PDGF and will facilitate the study of the role of PDGF in atherosclerosis and cancer.

One of the major problems in growth factor research has been to determine how the portion of the receptor inside the cell senses that the binding domain on the outside of the cell has interacted with PDGF. We have recently found that the transmission of the signal from the outside domain to the inside of the cell involves two major steps. First, when a receptor molecule binds PDGF, the receptor pairs with another identical receptor molecule to form a receptor dimer. Each of the two receptor molecules in the dimeric complex phosphorylates its partner, thereby modifying the partner. The phosphorylation reaction results in the addition of a phosphate group to the partner and is accomplished by a region of the receptor termed the kinase domain. We have designed and produced mutant receptors that have normal PDGF-binding domains but have defective kinase domains. When these mutants are introduced into cells that have normal receptors, a dimer (heterodimer) is formed between the normal and kinase-defective receptors. The normal and mutant receptors in the heterodimer complex cannot phosphorylate each other and cannot transmit the signals required to initiate cell growth. These experiments prove that formation of a dimer consisting of two normal recep-

tors is required for proper functioning and signal transduction of the receptor. By introducing mutant receptors into specific tissues of animals we hope to be able to block the function of the normal receptors and assess the role of PDGF in physiological processes and in disease states.

When the dimerized receptor is phosphorylated, the second major step in signal transduction occurs. The phosphorylated receptor physically binds to signaling molecules that are inside the cell. We have recently found that the interaction between the receptor and signaling molecules occurs at the phosphorylation sites on the receptor. Using information about the structures of these sites, we are designing ways to disrupt the interaction between the receptors and the signaling molecules. We are also using the receptors as a tool for discovering previously unidentified signaling molecules that mediate that proliferative response to PDGF. Recently we have identified one of these molecules, phosphatidylinositol 3-kinase (PI 3-kinase), that appears to play an important role in PDGF-stimulated proliferation. Other investigators have found that PI 3-kinase is also important in the cell transformation caused by some viruses that cause cancer in animals. Using the receptor as a probe, we recently purified the PI 3-kinase and cloned the gene for this signaling molecule. We hope that by studying the interaction of the receptor and PI 3-kinase we can understand an important set of reactions that are involved in regulating cell growth.

We recently have begun to study the fibroblast growth factors (FGF). These factors appear to play important roles in the earliest stages of embryogenesis and in angiogenesis (the formation of new blood vessels). The development of new vessels can be beneficial (e.g., in offsetting atherosclerotic narrowing of blood vessels in the heart) or deleterious (e.g., in the formation of new vessels that supply nutrients to tumors or in the vascular proliferation that occurs in the eyes of diabetic patients). We have identified some of the FGF receptors and are now examining their mechanisms of action. This work on FGF receptors is supported by a grant from the National Institutes of Health.

The long-range goal of these studies is to probe the role of growth factors in normal embryonic development, in tissue repair, and in proliferative diseases. Using the tools of molecular biology, cell biology, and protein chemistry, we and other research groups are identifying the factors, receptors, and regulatory molecules involved in these processes. Studies of the spatial and temporal distribution of the growth factors and receptors in normal and diseased tissues will provide insight into the function of these molecules. By understanding the molecular details of the protein-protein interactions involved in growth factor action, it may be possible to devise new therapeutic strategies to treat proliferative diseases.

Somatic Cell Gene Transfer

James M. Wilson, M.D., Ph.D.—Assistant Investigator
Dr. Wilson is also Associate Professor of Internal Medicine and Biological Chemistry at the University of Michigan Medical School. He received his undergraduate degree in chemistry from Albion College and his Ph.D. (biological chemistry) and M.D. degrees from the University of Michigan. He completed a residency and clinical fellowship at Massachusetts General Hospital, Boston, and conducted postdoctoral research with Richard Mulligan at the Whitehead Institute, Massachusetts Institute of Technology. Dr. Wilson is investigating ways to treat genetic diseases by correcting the basic defects.

GENE replacement therapies are being considered for the treatment of a variety of acquired and inherited diseases. These novel therapeutic modalities involve the transfer of genetic material into somatic tissues of affected individuals. The development of new therapies for cardiopulmonary diseases, based on gene transfer into hepatocytes and airway epithelial cells, is a major focus of my laboratory.

Familial hypercholesterolemia (FH) is an autosomal dominant disorder in humans that is an excellent model for developing gene replacement therapies of hyperlipidemic states. Patients with the homozygous deficient form of FH have severe hypercholesterolemia and suffer life-threatening coronary artery disease in childhood that is refractory to conventional medical therapies. The molecular basis of this disorder is a systemic deficiency in the receptor that degrades low-density lipoprotein (LDL), the primary carrier of cholesterol in the blood. Hepatocytes are the appropriate target for gene transfer in this disease, since the liver is the organ primarily responsible for LDL metabolism and cholesterol excretion.

We have used an animal model for FH—the Watanabe heritable hyperlipidemic (WHHL) rabbit—to develop different types of liver-directed gene replacement therapies. One approach is similar in concept to the well-described bone marrow–directed gene therapies. This *ex vivo* method involves isolating hepatocytes from a genetically deficient animal, transferring functional genetic material into the hepatocytes *in vitro*, and returning the corrected cells to the affected recipient. We have used recombinant retroviruses to transfer a functional gene for the human LDL receptor stably into hepatocytes derived from WHHL rabbits. Transplantation of these cells into WHHL rabbits leads to substantial decreases in serum cholesterol for over four months. A similar approach may be therapeutic in patients with homozygous FH.

An alternative and less invasive approach is to deliver a functional LDL receptor gene into the hepatocytes *in vivo*. We have developed an *in vivo* gene delivery system that is based on interactions with hepatocyte-specific receptors. Using this approach we can deliver reporter genes specifically to hepatocytes *in vivo* and obtain expression of the recombinant gene in liver for at least four months. Administration of a gene transfer substrate containing the LDL receptor gene into the circulation of WHHL rabbits led to significant reductions in the level of serum cholesterol. Hepatic overexpression of LDL receptor by gene transfer can potentially prevent the atherosclerotic disease in FH and other hyperlipidemic states. This work is also supported by a grant from the National Institutes of Health.

Cystic fibrosis (CF) is an inherited disease characterized by abnormal salt and water transport that leads to pathology within the pancreas and lung. CF is the most common lethal congenital disease among Caucasians, with a prevalence of 1 in 2,000 births. The primary defect in the lung appears to be the production of thick abnormal mucus that plugs the airways and leads to infections. The recent discovery of the gene that causes CF suggests a new therapeutic strategy in which normal functioning CF genes are directly introduced into pancreatic or respiratory cells of affected patients.

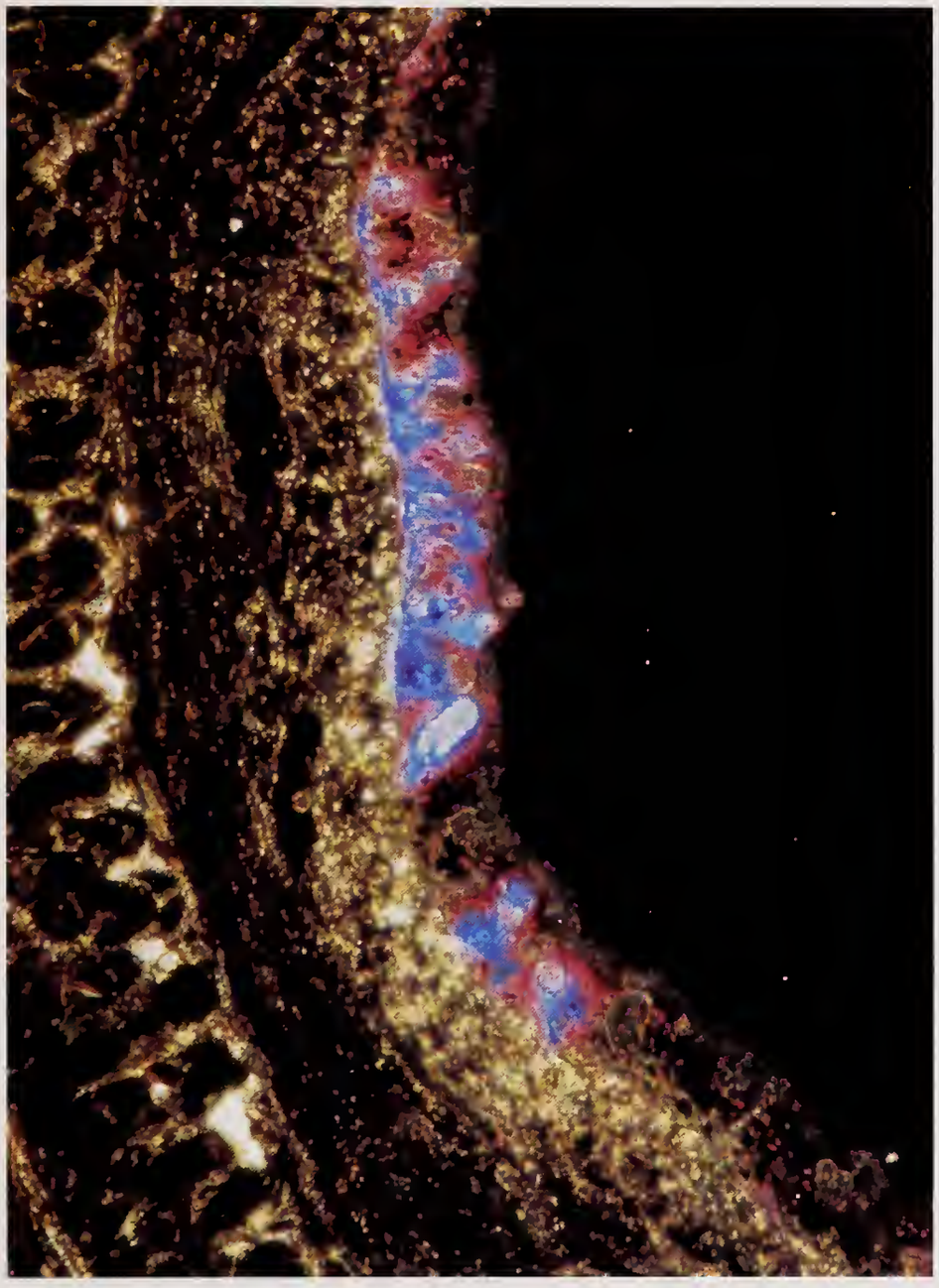
As a first step in the development of a genetic cure for CF, we have attempted to correct the physiological abnormality in cells from a CF patient by introducing into them a CF gene that encodes CFTR (cystic fibrosis transmembrane regulator), a normal functioning product. Replication-defective viruses were used to shuttle a normal CF gene into pancreatic cells from a patient with CF. Prior to gene transfer, the cells manifested the typical abnormalities associated with this disease (i.e., decreased transport of salt across the membrane). Following gene transfer, the cells regained normal regulation of salt transport. This demonstrates the feasibility of gene therapy in CF at a cellular level.

Rational development of approaches for reconstituting CFTR expression *in vivo* requires a definition of endogenous CFTR expression in the normal human airway as well as an understanding of the biology of the epithelial cells that line the

airway. Using antibody and RNA probes, we have characterized the distribution of CFTR expression in the human bronchus. The airway epithelial cells express low levels of CFTR. The predominant site of expression, however, is in the submucosal glands that are responsible for the production and delivery of mucus to the airway lumen. These findings suggest that effective gene therapies may require genetic reconstitution in both the epithelium and the submucosal glands.

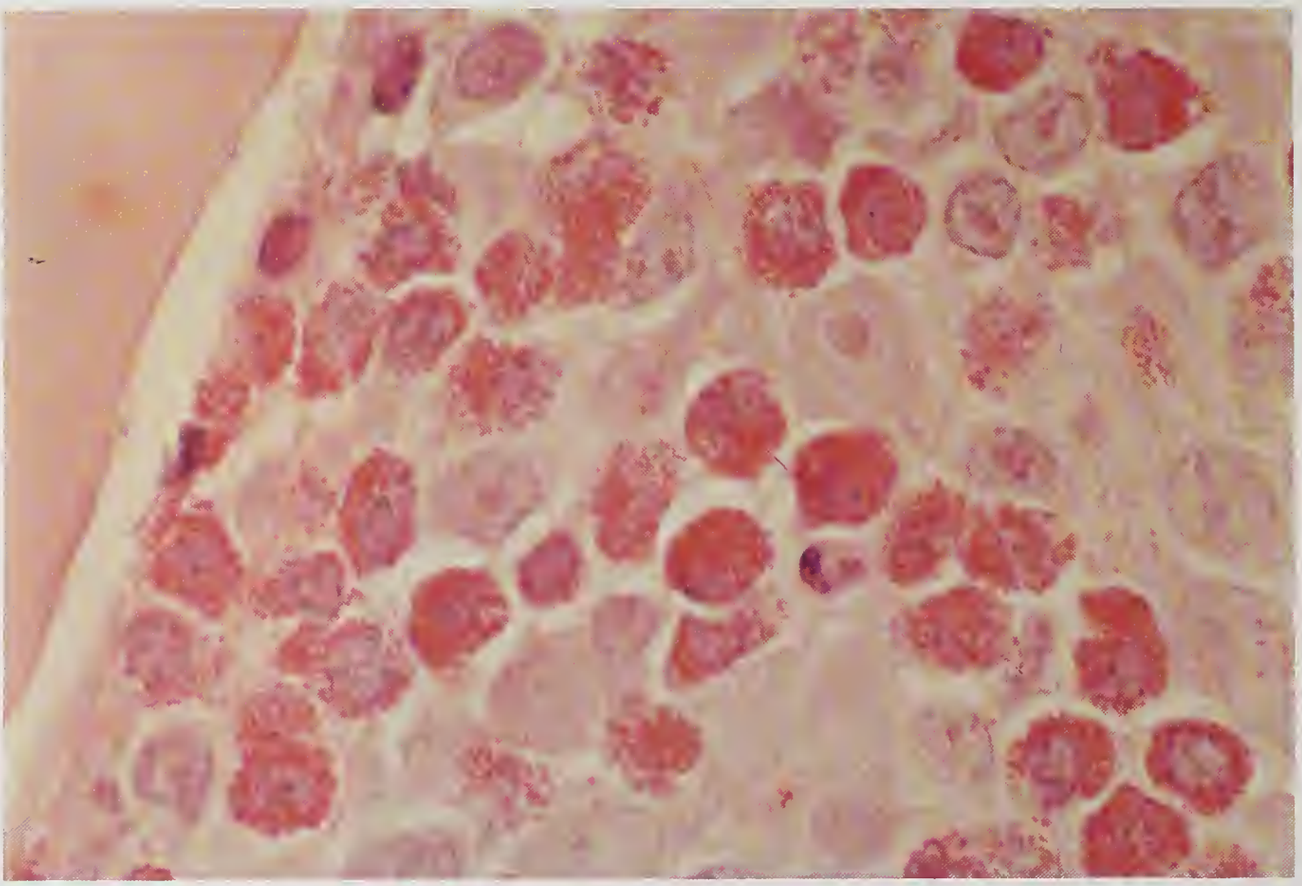
This work is also supported by the Cystic Fibrosis Foundation.

We have entered a new phase of the development of CF gene therapies in which approaches for delivering functional CF genes into airway cells *in vivo* are being designed and tested. Direct inhalation of the gene transfer substrates could provide a noninvasive way of delivering the genes to the appropriate cells. Several viral and nonviral approaches are being considered.



Model for gene therapy of the cystic fibrosis airway: human bronchial xenograft reconstituted with cells from a CF patient following retroviral infection with the β-galactosidase reporter gene. Darkfield photomicrograph shows blue cells following a histochemical stain for β-galactosidase.

Research and photograph by John Engelhardt in the laboratory of James Wilson.



Differentiation of human myeloid leukemia cells to eosinophils in the bone marrow of a mouse with severe combined immunodeficiency disease (SCID).

From Sawyers, C.L., Gishizky, M.L., Quan, S., Golde, D.W., and Witte, O.N. 1992. Blood *79:2089–2098.*

Normal and Abnormal Lymphocyte Growth Regulation

Owen N. Witte, M.D.—Investigator

Dr. Witte is also Professor of Microbiology and Molecular Genetics at the University of California, Los Angeles, holding the President's Chair in Developmental Immunology. He received his B.S. degree in microbiology from Cornell University and his M.D. degree from Stanford University, where he trained with Irving Weissman in the Medical Scientist Training Program. Dr. Witte completed postdoctoral training with David Baltimore at the Massachusetts Institute of Technology before joining the UCLA faculty. His honors include the Milken Family Medical Foundation Award in Basic Cancer Research and the Rosenthal Foundation Award of the American Association for Cancer Research.

OUR ability to resist a wide range of infectious agents depends on the normal function of the immune system. The humoral portion of this system is responsible for the production of specific antibody molecules from B lymphocytes. Too low a growth rate can result in immunodeficiency; too high a growth rate, in various types of leukemia or lymphoma. Our laboratory has concentrated on defining the growth control mechanisms that regulate the production of B lymphocytes.

The *ABL* Oncogene in Murine and Human Leukemias

The *ABL* oncogene was first isolated as the active genetic element of the Abelson murine leukemia virus. This agent is capable of causing a wide range of leukemias in mice. The biological properties of the *ABL* gene product depend on its activity as a tyrosine-specific kinase.

The human homologue of the *ABL* gene has now been strongly implicated in the pathogenesis of a family of human leukemias that harbor a specific cytogenetic abnormality. This is a chromosome translocation that uses mRNA splicing to join part of the *BCR* gene (of unknown function) from chromosome 22 to part of the *ABL* gene from chromosome 9, forming the so-called Philadelphia chromosome, or Ph1. The tyrosine kinase activity of the chimeric *BCR/ABL* gene product is evoked and strongly correlates with the transformation activity of the protein.

Two different forms of BCR/ABL protein can occur, depending on the precise position of the chromosomal breakpoints. In human chronic myelogenous leukemia, a larger protein product called P210 BCR/ABL is produced, and in the case of Ph1-positive acute lymphocytic leukemia, a P185 BCR/ABL protein product is commonly found. We have undertaken to determine the precise contribution of *BCR* sequences to the tyrosine kinase activity of the *ABL* segment and to the malignant potential of the gene product.

A variety of studies have now documented the specific role of BCR in the activation of the ABL tyrosine kinase. Site-directed mutagenesis has established that the BCR segment is essential for transformation by the chimeric oncogene and acts through the tyrosine kinase domain of ABL. We have recently described a new class of protein-protein interactions that regulate this activation. Within the first exon segment of BCR, there are two strong SH2 (SRC homologous domain 2) binding regions that can tightly bind to ABL. All previous SH2 interactions have been mediated by proteins containing phosphotyrosine. Interestingly, the BCR protein requires phosphoserine and phosphothreonine for strong binding, but not phosphotyrosine. BCR represents a new class of protein-protein interaction domains important in intracellular signaling.

Further analysis of BCR has shown that it also represents a new class of protein kinases. The first exon region of BCR contains a serine/threonine kinase activity that can be distinguished from the traditional protein kinase family by several criteria, including its containment of catalytic activity within a single exon, its reactivity with nucleotide analogues, and the role of essential cysteines within the nucleotide-binding cleft. Work from other laboratories shows that BCR also has a GTPase-activating protein domain at its carboxyl terminus that regulates the action of small guanine nucleotide–binding proteins. BCR appears to be a multifunctional protein. Its normal function remains to be determined.

The normal *ABL* oncogene products are expressed in many cell types, but their role in mammalian cell physiology is unknown. Gross structural changes can activate their oncogenic potential. It has been difficult to identify more-subtle mutations that might activate *ABL*, because the normal gene is toxic to most cell types when highly expressed. The precise mechanism of toxicity is not established, but probably relates to a cell cycle–blocking effect.

Full-length cDNA copies of the cellular *ABL* genes are cloned into a retroviral vector that has been modified to allow amplification in an acute transfection system. Retroviral particles are pro-

duced that can transmit the cellular *ABL* gene at high efficiency to a wide variety of cell types. Using this system, we have been able to select for transformed clones that harbor new classes of activating mutations in *ABL*.

One new class of mutants shows the loss of sequences downstream of the kinase domain. These naturally selected deletion mutants and molecular constructs created to mimic their structure are weak transforming alleles that require a complementing oncogene to induce fibroblast cells to grow in agar. Their most striking phenotypic characteristic is the very low activation of the tyrosine kinase activity of ABL. This is very different than all previous transformation alleles of *ABL*.

It has previously been shown that all forms of the *ABL* oncogene can synergize for their biological effects with members of the *MYC* gene family. We have recently tested the hypothesis that MYC is not only complementary to ABL transformation, but essential. Using a series of dominant negative mutants to inactivate the effective dosage of MYC in both fibroblast and hematopoietic cells, we have observed a dramatic decrease in the ability of BCR/ABL and viral forms of ABL to transform. The dominant negative mutants have no suppressive effect on other types of oncogenes, including *MOS*. Such essential combinations could be important considerations in devising strategies for the treatment of certain cancers.

Regulation of Stem and Progenitor Cell Growth

In collaboration with Naomi Rosenberg of Tufts Medical School, we have developed a system for retroviral infection of murine bone marrow stem cells with the *BCR/ABL* oncogene and reimplantation into syngeneic hosts. This procedure leads to tumors with the characteristics of human chronic myelogenous leukemia. Interestingly, animals infected with the P185 *BCR/ABL* forms show more-aggressive tumors that invade nonhematopoietic organs and show shorter latency than animals infected with the P210 form. This system should be valuable for analyzing new therapies directed at the *BCR/ABL* oncogene and for defining the growth regulation of primitive hematopoietic stem cells.

To elucidate the mechanisms that lead to deregulated growth of stem cells, we have developed an *in vitro* system to monitor the effects of BCR/ABL on multipotential stem cells. Marrow enriched for such cells was infected with P210 BCR/ABL expressing retrovirus and plated in agar in very low concentrations of growth factors like Steel factor and interleukin-3 (IL-3). These low concentrations are incapable of stimulating colony growth but could synergize with BCR/ABL.

A variety of colony types grew in this assay system, including multipotential cells capable of differentiating to form mast cells, macrophages, granulocytes, and pre–B cells. Cell lines of these different sublineages established from such multipotent colonies express a high level of the *BCR/ABL* gene product but are only subtly transformed in their growth factor requirements and nonmalignant when transferred to syngeneic animals.

Reliable test models for the growth of human myeloid leukemias are limited to a small subset of these leukemias that can be grown into continuous cell lines *in vitro*. We have recently established a reproducible procedure for the growth of freshly explanted human acute myelogenous leukemia and blast crisis specimens from chronic myelogenous leukemia patients. Cells from the bone marrow of such patients are implanted under the kidney capsule of the severe combined immunodeficient (SCID) mouse. At this site the human cells grow and then migrate to the bone marrow and peripheral blood in a pattern quite similar to that of the original human disease.

One long-range goal of our group is to develop effective *in vitro* culture techniques for the propagation and enrichment of stem and progenitor cells for different lineages. Previously we used the growth stimulatory properties of the *BCR/ABL* oncogene in concert with selected bone marrow–adherent stromal lines to grow clonal lines of B lymphoid progenitor cells that could repopulate the B cell lineage of immunodefective (SCID) mice. By modifying the culture conditions, we have now been able to cultivate such progenitor cells without the need for co-stimulation by the oncogene. These populations are very effective in reconstituting the B cell lineage *in vivo*.

We have used these progenitor cells to prepare several cDNA libraries in order to search for new members of the tyrosine kinase gene family and other regulatory genes that may be important in B cell development. A new member of the SRC family of tyrosine kinases has been identified that is specific to the B cell lineage.

Translational Regulation

Sandra L. Wolin, M.D., Ph.D.—Assistant Investigator

Dr. Wolin is also Assistant Professor of Cell Biology at Yale University School of Medicine. She received her undergraduate degree in biochemistry from Princeton University and her M.D. and Ph.D. degrees from Yale University, where she worked with Joan Steitz. Her postdoctoral research was done with Marc Kirschner and Peter Walter at the University of California, San Francisco, where she was a fellow of the Helen Hay Whitney Foundation and a Lucille P. Markey Scholar.

MY laboratory is particularly interested in the mechanisms that regulate the translation of messenger RNA (mRNA) into proteins. The information in mRNA is translated by a large RNA-protein complex, the ribosome. It has been known for some time that the ribosomes do not move along the mRNA at an even pace—that they pause at discrete places along the way. Why ribosomes pause is not well understood but is thought to be due either to rare codons or to secondary mRNA structures.

In certain mRNAs the pausing of ribosomes is thought to cause them to slip and lose their place, such that a different protein is now translated. This slipping is known as ribosome frameshifting. We have been investigating why ribosomes pause during translation, and we would like to understand how cells can use ribosome pausing to regulate the synthesis of particular proteins.

To address these questions, we use a method that allows us to determine the distribution of ribosomes on an mRNA with single-nucleotide precision. In this way we can obtain a detailed picture of the translation process. Using this assay, we have found that the ribosomes often pause directly over the initiation codon of the mRNA. This pausing by fully assembled ribosomes appears to represent a slow step in eukaryotic protein initiation that had not been previously detected.

By performing translation reactions in the presence of inhibitors that block distinct steps in polypeptide initiation and then examining the position of the resulting paused ribosomes, we have narrowed down this major slow step to one of two points in the initiation pathway. We are continuing to characterize this intermediate in translation initiation and are also investigating the possibility that this slow step in protein synthesis may be a point at which cells regulate translation of particular mRNAs.

We are also interested in understanding how the availability of particular tRNA molecules (intermediates in protein synthesis) affects ribosome pausing during elongation of the nascent protein. To study this question, we have prepared translation extracts in which the tRNA molecules have been removed. By adding back different tRNA populations, we can now manipulate the levels of individual tRNAs. In this way we hope to determine how different tRNA species contribute to ribosome pausing and frameshifting.

We are also studying how the attachment of ribosomes to the endoplasmic reticulum membrane affects their movement along mRNAs encoding secreted proteins. The synthesis of these proteins designed for export outside the cell begins when ribosomes, free in the cytoplasm, initiate translation on the mRNAs. A common feature of these secretory proteins is the presence of a signal peptide, usually an amino-terminal extension of 15–30 amino acids. A signal-recognition particle (SRP), a small cytoplasmic ribonucleoprotein, binds to the signal sequence emerging from the ribosome and arrests elongation transiently. This translational arrest is released after the ribosome-SRP complex interacts with a specific component of the endoplasmic reticulum membrane, the SRP receptor.

Concomitant with the resumption of protein synthesis, translocation of the nascent polypeptide begins. All subsequent translation is carried out by ribosomes that are attached to the endoplasmic reticulum membrane.

We have found that the point at which membrane insertion of the nascent polypeptide begins is distinct for different proteins, which may reflect differences in the size and structures of individual signal peptides. It also appears that certain features of mRNA structure that cause ribosomes to pause during translation in solution do not always result in their pausing when attached to microsomal membranes. Thus it appears likely that mRNA secondary structure can be altered by attachment through ribosomes to these membranes.

We are currently determining whether ribosomes, following termination, are able to reinitiate translation on microsomal membranes. If so, it would provide a molecular explanation for the "circular polysomes" that have long been observed attached to the endoplasmic reticulum.

Molecular Genetics and Studies Toward Gene Therapy for Metabolic Disorders

Savio L. C. Woo, Ph.D.—Investigator

Dr. Woo is also Professor in the Department of Cell Biology and Institute for Molecular Genetics at Baylor College of Medicine. He obtained his undergraduate degree in chemistry from Loyola College, Montreal, and his Ph.D. degree in biochemistry from the University of Washington, where he worked with Earl Davie. His postdoctoral research was done at the University of British Columbia, Vancouver, in the Division of Neurological Sciences.

A major focus in my laboratory has been the analysis of human metabolic disorders at the molecular level. Phenylketonuria (PKU), the disease under investigation, is an inborn error in amino acid metabolism that causes severe and permanent mental retardation in untreated children. The condition is caused by defects in the liver enzyme phenylalanine hydroxylase (PAH) and is transmitted from the parents to both boys and girls. In the United States it affects 300–400 newborns a year, and 1 of every 50 individuals is an asymptomatic carrier of the disease trait. A second area under intense investigation is technology development for the cure of these genetic disorders by somatic gene therapy. The diseases being studied here include PKU and hemophilia B, which is transmitted from asymptomatic carrier mothers to their sons.

Prenatal Diagnosis for Phenylketonuria

Our laboratory has isolated the human PAH gene by molecular cloning and used the cloned gene to analyze cellular DNA of normal and affected individuals. Extensive benign variations in this gene were discovered and used to trace the transmission of individual PAH genes from the parents to the children in PKU families. A fetus inheriting the same PAH genes as an affected sibling will have PKU. This has led to a prenatal diagnosis procedure for PKU in families with previously affected children, and the procedure has been adopted in the United States and other western European countries.

Prognosis by Gene Analysis

There is a wide range of severity of clinical symptoms among PKU patients, and their treatment depends on measurement of the phenylalanine level in blood, which is often highly variable. The development of an independent method for the determination of prognosis is important for proper management. We have analyzed the mutations in the PAH gene of a large number of PKU patients who have been closely supervised for the past 20 years. The severity of their clinical conditions is primarily dependent on the inheritance of mutations in their PAH genes that are either totally or partially defective. In the future this correlation will allow physicians to prescribe proper medical treatment after simple gene analysis.

Population Dynamics

A number of these PKU genes have distinctive distributions in the European continent. One is very prevalent in eastern Europe, and its frequency decreases in a gradual fashion from east to west. These results suggested that the mutation occurred in eastern Europe some time ago and was then spread throughout the European continent by migration of people in prehistoric times. Two other prevalent PKU genes are very frequent in either northern or southern Europe but less frequent in the neighboring countries. When similar analysis was performed in Israel and China, independent centers of major PKU mutations were also discovered. These results strongly suggest that multiple PKU mutations occurred independently in various regions of the world and then spread into neighboring areas in Europe and Asia.

Somatic Gene Therapy

The other major goal of our laboratory is to explore the potential for somatic gene therapy of genetic disorders. The PAH gene has been inserted into the genome of an incapacitated virus. The recombinant viruses are able to transduce mammalian cells and incorporate themselves into the genome of the host cells, but they are no longer able to produce new virus to continue the infection process. The recombinant viruses were used to transduce cultured rodent hepatoma cells and normal liver cells, thereby conferring upon them the ability to synthesize the corresponding human enzyme.

These results have led to the development of hepatocyte transplantation technologies in laboratory animals. A variety of inert substances were used as support for mouse liver cells prior to transplantation, but the transplanted hepatocytes

lived for only a few weeks in the recipient animals. When the hepatocytes were returned to the liver by direct injection into the portal vein or the spleen of recipient mice, however, the cells migrated to the liver and incorporated themselves into the parenchyma. They not only lived for as long as the mice did but also continued to function as liver cells in the transplanted animals. Using liver cells isolated from a mouse strain that is deficient in hepatic PAH activity, we demonstrated that enzymatic activity was reconstituted in these cells after retroviral-mediated gene transfer. The PKU mouse model will be critically important to test the efficacy of our hepatic gene transfer and hepatocyte transplantation protocols for the correction of PKU in the future.

To develop a larger animal model for somatic gene therapy, with technologies that may be directly applied to human patients in the future, we selected a colony of hemophilic dogs. Methods were developed to obtain a liver lobe from normal dogs by partial hepatectomy and to disperse the hepatocytes into single cells in culture. Instead of a few million cells (which we obtained from the mouse), several billion cells were obtained from a single canine liver lobe. The isolated hepatocytes were transplanted back into the same animal by direct injection into the spleen. Alternatively, a catheter was inserted into the splenic vein with a subcutaneous port to permit direct external injection of hepatocytes into the port. It was observed that a billion hepatocytes can be easily transplanted, and greater than half of these cells migrate to the liver and survive in the animal for a minimum of four months. Using a recombinant retroviral vector containing the human coagulation factor IX gene to transduce normal canine hepatocytes, we also observed high levels of human factor IX protein in the culture media. With these technological developments, we shall attempt to correct the genetic deficiency in hemophilia B dogs by somatic gene therapy.

Prenatal diagnosis of PKU and disease prognosis by gene analysis were supported by a grant from the National Institutes of Health. Somatic gene therapy research on the PKU mouse and hemophilia B dog models is also supported in part by a grant from the National Institutes of Health.

Paracrine Control of Blood Vessel Function: Role of the Endothelins

Masashi Yanagisawa, M.D., Ph.D.—Associate Investigator
Dr. Yanagisawa is also Associate Professor of Molecular Genetics at the University of Texas Southwestern Medical Center at Dallas. He received his M.D. and Ph.D. degrees from the University of Tsukuba, Japan, where he worked with Tomoh Masaki. Before moving to Dallas, he was Assistant Professor of Pharmacology at Kyoto University, Japan.

DISORDERS of blood vessels, including all forms of heart attack and stroke, represent the most frequent disease cause of death in developed countries. The inner surface of blood vessel walls is lined with a thin monolayer of flat cells called vascular endothelium. These cells cover a total surface area of nearly 700 m^2 throughout the human body. The endothelium is also unique in that it is the only tissue that has direct physical contact with circulating blood under healthy conditions. Nevertheless, until a little over 10 years ago, the endothelial cells were considered to be merely a "bio-inert dialysis bag." In other words, they just kept blood flowing smoothly without unwanted clotting and allowed nutritional components and metabolites to pass freely between the blood and interstitial space.

Evidence accumulated recently, however, indicates that the endothelium plays much more complex roles in many different facets of physiology and pathology. The endothelial cells can respond to both chemical (circulating and local hormones) and mechanical (local blood flow and pressure) information carried in the circulating blood.

In response to these stimuli, these cells trigger remodeling and *de novo* formation of blood vessels by secreting growth factors and migrating into adjacent tissues. They also regulate leukocyte infiltration and lymphocyte homing by expressing a variety of cell adhesion molecules. They integrate blood coagulation, fibrinolysis, and platelet function by producing various pro- and anticoagulant factors. They control vascular permeability by actively transcytosing and metabolizing plasma components. Finally, they regulate blood pressure and local blood flow by both activating and inactivating circulating vasoactive factors and, more importantly, by generating an array of vasoactive molecules. Thus we now recognize the vascular endothelium as an active and dynamic transducer that senses and interprets blood-borne signals.

Central to these functions of the endothelium is local communication of the endothelial cells with vascular smooth muscle cells underneath. As mentioned above, the endothelial cells send out an array of physiological and pathological signals toward the smooth muscle cells by means of the endothelium-derived vasoactive factors. Among those are prostacyclin and nitric oxide (also called endothelin-derived relaxing factor), both strong vasodilators.

This background led us in 1988 to identify endothelin-1 (ET-1), a novel vasoconstrictor produced by vascular endothelial cells. ET-1 is a small peptide consisting of 21 amino acid residues wound into a rigid structure by two sets of intrachain disulfide bridges. It is the most potent vasoactive molecule known, causing a strong and extremely sustained constriction of all blood vessels both *in vitro* and *in vivo*.

Our subsequent studies demonstrated in humans and other mammals three endothelin-related genes that encode, besides ET-1, two additional isopeptides of the endothelin family called ET-2 and -3. Endothelin isopeptides are expressed with distinct distribution patterns in many mammalian tissues. While the endothelial cells are the most abundant source of ET-1, one or more of the three isopeptides are expressed widely in other tissues such as brain, lung, and kidney.

The production of endothelins is regulated in both directions. In endothelial cells, it is upregulated by various chemical stimuli, including classical vasoactive hormones (epinephrine, angiotensin II, and vasopressin), products from coagulation/platelet activation (thrombin and transforming growth factor-β), factors implicated in septic shock (bacterial endotoxin, interleukin-1, and tumor necrosis factor), oxidized low-density lipoprotein, etc. ET-1 production is also augmented by mechanical stimuli such as stretch and fluid-dynamical shear stress. In contrast, certain vasodilators (nitric oxide and atrial natriuretic peptides) attenuate the ET-1 production via increase of intracellular cGMP levels.

Like many other peptide hormones and neuropeptides, endothelins are processed from the corresponding larger precursor proteins (pre-pro-proteins). However, biologically active 21-

residue endothelins are produced via a formerly unknown type of proteolytic processing. Biologically inactive intermediates of approximately 40 amino acids, called big endothelins, are first cut out from the prepro-proteins. Their carboxyl-terminal half is then sheared off to produce the amino-terminal active peptides.

This unusual endoproteolytic activation is catalyzed by a novel membrane-bound metalloprotease called endothelin-converting enzyme. This unique enzyme is sensitive to the metalloprotease inhibitor phosphoramidon and is distinct from any other protease known. Since big ET-1 is at least 100–200 times less active than mature ET-1 in constricting vascular smooth muscle strips, the converting enzyme is essential in producing the potent vasoconstrictive agent. Therefore the enzyme could also be an important target for pharmacological intervention in the endothelin system. If one could develop an inhibitor of endothelin-converting enzyme, it could possibly be used as a novel class of vasodilatory drug.

Apart from their potent and long-lasting vasoconstrictor/pressor activities, endothelins possess a wide spectrum of nonvascular actions in various tissues. Furthermore, two distinct subtypes of endothelin receptors (called ET_A and ET_B) with different selectivities to the three isopeptides have been pharmacologically demonstrated and molecularly cloned by us and other researchers. In keeping with the wide spectrum of endothelin action, the two receptors are expressed in a variety of vascular and nonvascular tissues.

The observation that many endothelin-producing tissues also express one or more subtypes of endothelin receptors suggests the importance of the peptide family as locally acting mediators (paracrine and/or autocrine) rather than circulating agents. The idea is further supported by several lines of evidence. For example, clearance of endothelins from the circulation is extremely rapid, with an initial half-life of less than 2–3 min; plasma concentration of immunoreactive ET-1 is well below the threshold concentration for pharmacological actions; and circulating endothelins are capable of inducing the release of vasodilator substances such as nitric oxide and prostacyclin via the receptors on the endothelial cells, thereby limiting their own vasoconstrictor actions.

Increased plasma concentrations of ET-1 have been reported in patients with various disease states, including vasospasm and hypertension, where abnormal vasomotor function is implicated. Moreover, in animal models of certain vascular disorders, such as acute myocardial infarction, cerebral vasospasm, and acute ischemic renal failure, treatment with antiendothelin neutralizing antibody significantly ameliorates the abnormal vasoconstriction seen in these conditions.

It is hoped that further insight into the physiological and pathobiological roles of this complex system of peptidic mediators will be gained in the near future with the development of endothelin receptor antagonists, inhibitors of endothelin-converting enzyme, and mice deficient in endothelin/endothelin receptors. This may lead to a new level of understanding of how cardiovascular homeostasis is maintained via local communication between cells of blood vessel walls in health and disease.

Mechanism of Phototransduction in Retinal Rods and Cones

King-Wai Yau, Ph.D.—Investigator

Dr. Yau is also Professor of Neuroscience at the Johns Hopkins University School of Medicine. He received an A.B. degree in physics from Princeton University and a Ph.D. degree in neurobiology from Harvard University. He did postdoctoral research at Stanford University with Denis Baylor and at Cambridge University, England, with Alan Hodgkin. For six years thereafter, he was on the faculty at the Department of Physiology and Biophysics of the University of Texas Medical Branch at Galveston. He has received the Rank Prize in Optoelectronics from the Rank Prize Funds, England.

VISION begins in the rods and cones of the retina, where light is absorbed and transduced into a neural signal consisting of an electrical hyperpolarization at the photoreceptor membrane. This signal is relayed to second-order neurons in the retina through a modulation of the release of synaptic transmitter at the photoreceptor's terminal. In darkness the transmitter is released at a high rate, and in light the membrane hyperpolarization reduces the release in a graded fashion. This modulation of synaptic transmitter release can lead to a hyperpolarizing or depolarizing response to light in a second-order neuron, depending on the polarity of a given synapse.

The phototransduction process—the way the hyperpolarizing response to light is generated in the receptors—is as follows. In darkness an ionic conductance in the plasma membrane of the receptor's outer segment (the part of the cell that contains the visual pigment) is kept open by the cyclic nucleotide guanosine 3′:5′-cyclic monophosphate (cGMP), letting both Na^+ and Ca^{2+} into the cell. This "dark" current depolarizes the cell and causes the steady release of synaptic transmitter described above.

Light activates the following reaction cascade: light → photoisomerization of visual pigment → G protein activation → cGMP phosphodiesterase stimulation → cGMP hydrolysis. As a result, the cGMP level falls in the outer segment, causing the ionic conductance to close and leading sequentially to membrane hyperpolarization and the reduction of synaptic transmitter release. This phototransduction scheme applies to both rods and cones, with only quantitative differences between the two types of receptors.

One consequence of the conductance closure in the light is that the Ca^{2+} influx stops. The resulting imbalance between influx and efflux leads to a decrease in the intracellular free Ca^{2+} concentration. This Ca^{2+} decrease reduces a tonic inhibition exerted by Ca^{2+} on the cGMP-synthesizing enzyme guanylate cyclase and causes an increase in the synthesis of cGMP in the light. Thus Ca^{2+} mediates a negative feedback control on the light-activated cGMP hydrolysis, and this feedback should be a candidate mechanism underlying the well-known phenomenon of background light adaptation in photoreceptors. Indeed, we have found that this adaptation essentially disappears upon removing the feedback experimentally by eliminating the Ca^{2+} influx and efflux.

The effect of Ca^{2+} on rod guanylate cyclase has been studied by others in biochemical experiments *in vitro*. This effect is now known to involve recoverin, a novel Ca^{2+}-binding protein that activates guanylate cyclase at low Ca^{2+} concentrations but loses this ability when Ca^{2+} is bound to it. One drawback of the biochemical experiments is that unphysiological ionic concentrations (e.g., very high Mg^{2+}) were used to measure the cyclase activity. We have now studied this Ca^{2+} modulation of the cyclase in more physiological conditions, by recording the cGMP-activated current from an isolated, open-ended rod outer segment with a suction pipette while dialyzing the interior of the outer segment with different Ca^{2+} concentrations.

The Ca^{2+} effect on the guanylate cyclase could be derived from the magnitude of the cGMP-activated current. We have found that the cyclase activity is very sensitive to the free Ca^{2+} concentration, with its maximum activity being approximately halved at 100 nM Ca^{2+}. This is similar to biochemical measurements. The cyclase dependence on Ca^{2+} shows a Hill coefficient of approximately 1.5, which is lower than the value of around 4 in biochemical measurements. The cyclase activity becomes relatively insensitive to Ca^{2+} at a GTP concentration of greater than 1 mM, suggesting that the effect of Ca^{2+}, acting through recoverin, may primarily be to reduce the affinity of the enzyme for its substrate. Further experiments on this problem are in progress.

Another problem we are working on is a molecular characterization of the cGMP-activated conductance mediating phototransduction. The conductance now appears to belong to a family of cyclic nucleotide–gated channels that includes the cGMP-gated channels in retinal rods and cones, as well as the cGMP/cAMP-gated channel

mediating olfactory transduction in olfactory cilia. These channels show both similarities and differences. In particular, the rod and olfactory channels have similar ion-permeation characteristics, but the olfactory channel shows a 30-fold higher affinity for cyclic nucleotides. To understand the molecular determinant for this difference, we, in collaboration with Randall Reed (HHMI, Johns Hopkins University), have constructed chimeras between the rod and the olfactory channels and tested for modifications in function. The difference in amino acid residues at the putative cyclic nucleotide–binding site on the two channel molecules has little to do with their difference in affinity for cyclic nucleotides. Other regions of the molecule must therefore be involved, possibly through steric or allosteric interactions in the folded molecule. Identification of these regions is still in progress.

Separately, several independent clones have been isolated from a human retinal cDNA library based on structural homology to the published sequence of the bovine rod channel. One of these has been successfully expressed in human 293 cells and found to have properties identical to the bovine rod channel. Thus this clone most probably encodes for the human rod channel. The other clones have yet to be expressed functionally in a cell line. Based on the presence of a putative cyclic nucleotide–binding domain as well as other features in their nucleotide sequences, however, they probably also encode for cyclic nucleotide–gated channels. Antipeptide antibodies are being made to identify the locations of the encoded proteins in the retina. Continuing efforts are also being made to express these clones functionally. Cloning at the genomic level is being carried out to provide further clarification.

Part of the above work is supported by a grant from the National Institutes of Health.

Molecular Mechanisms of Ion Channel Function

Gary Yellen, Ph.D.—Assistant Investigator

Dr. Yellen is also Assistant Professor of Neuroscience and Biophysics at the Johns Hopkins University School of Medicine. He received his undergraduate degree in biochemical sciences from Harvard College and his Ph.D. degree in physiology from Yale University, where he studied with Charles Stevens. Dr. Yellen did his postdoctoral research on ion channel physiology as a Life Sciences Research Foundation postdoctoral fellow at Brandeis University, where he worked with Christopher Miller.

ALL electrical signaling in the nervous system is controlled by ion channels, a class of membrane proteins that form pores through the membrane. Charged ions such as sodium, potassium, and calcium pass through ion channels and carry an electrical current. The channels themselves are regulated, so that the pores are only open when the proper chemical or electrical signal is present, and only certain ions can pass through a particular kind of channel. By understanding how channels open and close and how they are regulated, we define the repertoire of molecular changes used by neurons when they signal, sense, and learn.

Ion channels, like other membrane proteins, have resisted standard biochemical and structural analysis. Their structure has only recently begun to be elucidated by a combination of protein chemistry and molecular biology. On the other hand, we have detailed knowledge of the functioning of ion channels. Because each ion channel catalyzes the transport of millions of ions per second, we can measure electrically the current carried by just a single-channel protein molecule. This technique of single-channel recording has allowed us to make a detailed model for the conformational changes between open and closed states induced by chemical ligands and changes in voltage, but we still have no knowledge of the protein structures that underlie these conformational changes.

My laboratory uses a combination of high-resolution functional analysis (by single-channel recording) and direct manipulation of the structure of the channel protein. Site-directed mutagenesis allows us to modify any amino acid in a protein for which we have the cloned genetic material. Rather than modifying the protein directly, we change the DNA sequence and then inject the modified messenger RNA into immature frog eggs (oocytes), which manufacture the modified protein. This method allows us to test specific theories about which parts of the channel protein are important for specific functional features.

We have applied this combination of strategies to voltage-activated potassium channels, which participate in electrical signaling. By systematic mutagenesis, we have identified the region of the potassium channel protein that lines the pore through which ions cross the membrane. We found specific amino acid residues in the protein sequence that control the sensitivity of the channel to tetraethylammonium, an organic ion that can block current through the channel. Natural variation of one of these amino acids explains the differences in drug sensitivity between potassium channels in different organs or species. Amino acids in this region of the protein can also affect the rate at which ions are conducted through the pore. These discoveries put us in a position to discover the basic mechanisms of ion selectivity and channel gating at the level of individual amino acids.

We have also used recording from single potassium channels to demonstrate that one of the mechanisms by which these channels open and close is a simple occlusion of the pore by part of the channel protein. Earlier work established that the pore could be directly blocked or occluded by internal organic ions; we have established that the natural gating occurs by a very similar mechanism involving a tethered blocking particle. The most direct demonstration of this is that potassium ions passing through the pore from one side can clear the tethered blocking particle from the opposite side.

Further work in progress to determine the structural basis for potassium channel gating includes introducing chemically reactive cysteine residues at specific locations in the protein sequence. Channel proteins with cysteines inserted at critical locations should show a specific change in function when treated with reagents that modify the cysteine side chain. The reactivity of these side chains will depend on both their location in the channel protein and the specific conformational state of the protein at the time of reaction.

We have also used the site-directed strategy to study acetylcholine-activated cation channels, which convert neurochemical signals into electrical signals at synapses. We have changed amino

acids in the region of the protein that binds acetylcholine and identified specific residues that play a critical role in binding and signal transduction by acetylcholine. These studies are teaching us more about the molecular basis of drug recognition and of signal transduction in this protein.

Dr. Yellen is now Associate Professor of Neurobiology at Harvard Medical School.

Drosophila *Behavior and Neuromuscular Development*

Michael W. Young, Ph.D.—Investigator

Dr. Young is also Professor of Genetics at the Rockefeller University. He received his B.A. degree in biology from the University of Texas, Austin. Staying on to work at the university with Burke Judd, he earned his Ph.D degree for genetic and cytological studies of Drosophila *chromosome structure. Dr. Young did postdoctoral work in biochemistry at Stanford University Medical School with David Hogness. He moved to Rockefeller as a fellow of the Andre and Bella Meyer Foundation.*

A biological clock, composed of a few thousand cells within the mammalian brain, controls timing of daily behaviors such as sleep with an accuracy of minutes. Chemical and electrical rhythms have been detected in these mammalian pacemaker tissues. Still, little is known about the underlying biochemistry used to calculate time.

The genes and proteins central to biological timing are beginning to be recovered and characterized in a simpler model organism, the fruit fly *Drosophila*. The best-studied gene in the *Drosophila* clock system has been named *per* (*period*). Several mutant forms of the gene have been recovered that affect the pace of the insect's clock and certain aspects of cell physiology.

In the per^{l} mutant, circadian locomotor activity rhythms have a long period of 30 rather than 24 hours. For the mutant per^{s}, daily cycles have a shortened, 19-hour period. Mutants with no daily rhythms are designated per^{0}. Corresponding changes in cycle time are found for a high-frequency rhythm—a courtship song produced in males by pulses of wing beating: an 80-second song (instead of the normal 55 seconds) for per^{l}, 40 seconds for per^{s}, and song arrhythmicity for per^{0}. Also, the mutations change the period of a daily oscillation in *per* transcription, which may be important for establishing rhythmic behavior. Finally, for at least some tissues, the mutants appear to alter conductance of specialized channels (gap junctions) between cells.

The molecular changes associated with the mutations have been established. The per^{0} mutant cannot express a full-length protein. On the other hand, per^{s} and per^{l} make *per* proteins, but these are altered by substitution of a single, different amino acid. Comparable changes in cycle time can also be effected by altering the amount of *per* protein the fly produces. For example, microinjection of a gene that underproduces the *per* protein 20-fold induces 40-hour daily rhythms. From these results it has been suggested that the per^{s} and per^{l} mutations generate, respectively, hyper- and hypoactive proteins.

In a recent effort to understand how changes in protein structure can affect *per* activity, genes carrying new mutations, produced *in vitro*, were reintroduced into the fly by microinjection. It was found that mutations changing the structure of a certain segment of about 20 amino acids predominantly confer short-period (per^{s}-like) rhythmicity. Apparently the mutations identify a region of the *per* protein that regulates the activity level in normal flies.

A variety of experiments have demonstrated that the *per* protein acts in the nervous system to control daily and circaminuten rhythms. We have become interested in tracking the development of the fly's clock in an effort to determine when it begins to run, whether it requires signals from outside the organism to start, and where the first cells expressing *per* arise and develop. We have learned that *Drosophila* reared in constant environmental conditions spontaneously start their clocks only hours after formation of the embryo. Evidence of a running clock is first seen after completion of the embryonic nervous system and just following cessation of high levels of *per* expression in certain neural cells.

Until recently, only the *per* locus was known to be essential for production of biological rhythms in the fruit fly. Genetic screens for rhythm mutations have led to the discovery of additional, indispensable genes. Of special interest is a new mutation found on the second chromosome (*per* maps to the X chromosome). The new mutation, like per^{0}, renders flies arrhythmic, and in molecular studies appears to block the circadian rhythm in *per* transcription. Thus the newly recognized gene may be required for *per* to function.

Development of Skin, Muscle, and Nerve

In the embryo the nervous system and skin are derived from a common set of cells, the ectoderm. Each of these cells must choose a fate, and in certain *Drosophila* mutants the choice goes awry. For one set of mutants known as "neurogenic," the capacity to choose skin has been lost and only nerve is formed. We believe the mutants have lost the ability to form a set of signals that act early in development, so we are using the muta-

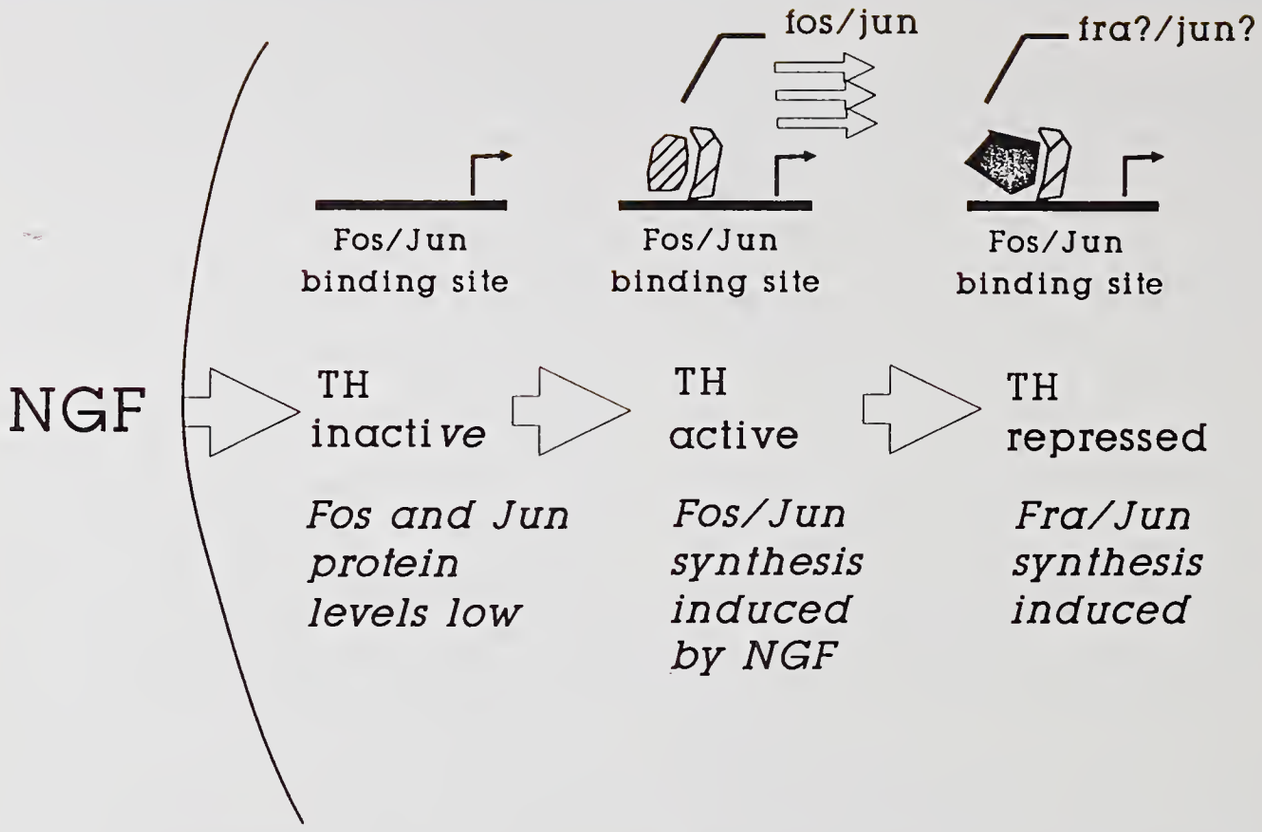
tions to mark, isolate, and characterize the products of these genes and putative developmental signals. Neurogenic mutations have been recovered at seven genetic loci: *Notch*, *big brain*, *mastermind*, *neuralized*, *Delta*, *almondex*, and *Enhancer of split*.

Notch has been characterized more completely than any other gene in the neurogenic group. It produces a very large, 2,700–amino acid protein, predominantly composed of an uninterrupted array of 36 copies of a hormone-like molecule, a relative of epidermal growth factor. The *Notch* protein spans the cellular membrane, with the 36 hormone copies exposed to its neighbors. We surmise that the neighbors read that signal and, in return, tell the *Notch*-bearing cell to come up with the correct allocations of skin and nerve. Since the entire string of hormone copies is tethered to the cell's surface, signaling between cells must be intimate. Only cells that can touch each other could communicate through such a protein.

From work with temperature-sensitive mutations, we know that *Notch* proteins are used to instruct development throughout embryonic, larval, and pupal life. Mutations altering the structures of individual hormone-like elements of the *Notch* protein have been examined to determine the role each plays in early and late development. Of the 36 hormone repeats, no two are identical, and changes in different hormone elements produce different developmental abnormalities. This must mean that alternate parts of the *Notch* hormone string are read as development unfolds. In part these specificities could allow a cell to talk to changing neighbors from the time of cell birth to differentiation into adult tissue.

What do signals from these genes tell a cell to do? For several years gene action at *Notch*, *Delta*, *big brain*, *almondex*, *neuralized*, *mastermind*, and *Enhancer of split* have been assumed to stimulate an undifferentiated ectodermal cell to develop as an epidermal cell. The genes are thought to provide a series of epidermalizing signals during cell differentiation, with loss of function generating a nerve cell. New work in our laboratory shows this simple picture to be inaccurate. *Notch* proteins have now been found in cells giving rise to embryonic muscle. In *Notch* mutants, strong effects on muscle development are seen, with increased numbers of some muscle cell types generated, probably at the expense of others. Thus parallel changes in muscle, skin, and nerve development take place in *Notch* mutants. Of most significance, comparable effects on muscle development are seen with mutations of *Delta*, *big brain*, *mastermind*, *almondex*, *Enhancer of split*, and *neuralized*.

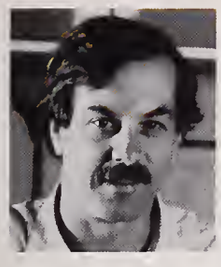
We have learned three things from these studies: 1) the developmental fates of many cell types are switched in neurogenic mutants; 2) the genes must provide differentiation signals that cells composing different germ layers can read, with no apparent overlap in the final developmental fates; and 3) the genes defined by the neurogenic mutations probably work together to form a single developmental pathway, which generates a common differentiation signal in all cells affected in the mutants, since any developmental anomaly caused by loss of one gene in the group predicts a comparable developmental change upon loss of any other gene in the series.



Model for a mechanism of control over a gene that may play a role in Parkinson's disease. The polypeptide hormone NGF (nerve growth factor) binds to receptors on the surface of nerve cells (neurons) and induces expression of the gene encoding tyrosine hydroxylase (TH). This enzyme carries out a critical step in the synthesis of the neurotransmitter dopamine, whose release by neurons transmits signals to neighboring cells. Thus NGF may indirectly control the neuron's signaling capacity. In the model, NGF induces synthesis of members of the Fos family of proteins, which form complexes with Jun family proteins. These complexes bind to a regulatory element of the TH gene and stimulate TH expression. Later, proteins related to Fos—Fra's—are expressed and repress TH gene activity. A deficiency in dopamine production can lead to Parkinson's disease and other neurological disorders.

Research of Elena Gizang-Ginsberg and Edward Ziff.

Control of Transcription by Transmembrane Signals



Edward B. Ziff, Ph.D—Investigator

Dr. Ziff is also Professor of Biochemistry at New York University Medical Center. He received his B.A. degree in chemistry from Columbia University and the Ph.D degree in biochemistry from Princeton University. He then studied DNA structure with Fred Sanger at the MRC Laboratory of Molecular Biology in Cambridge. He later conducted research on DNA tumor viruses at the Imperial Cancer Research Fund Laboratory, London, and in the Department of Molecular Cell Biology at the Rockefeller University. He later joined New York University Medical Center and began the study of cellular mechanisms that control proliferation and differentiation.

THE remarkable process of development requires that a fertilized egg undergo many rounds of cell division with accompanying differentiation to generate specialized cell types that ultimately form the mature organism. For development to proceed normally, just the right number and types of cells must be available at each stage. It follows that any cell's decision to divide and/or differentiate must be carefully regulated. These two critical processes, proliferation and differentiation, are also controlled in the adult—for example, during the maintenance of tissues and in wound healing.

The decision of a cell to divide or to express a specialized function is often determined by signals from its environment. Prominent among the agents that convey such signals are the growth factors, which are polypeptides synthesized and secreted by cells. Our laboratory studies the molecular mechanisms by which growth factors and other transmembrane signaling agents exert their effects on cell proliferation and differentiation. The transforming genes of DNA tumor viruses, such as adenovirus, often modify these programs and are therefore useful for dissecting the growth regulatory pathways.

Growth factors transmit signals to cells by binding to specific receptor proteins, which span the cell's plasma membrane and induce second messenger signals in the cytoplasm. The latter signals have a multitude of targets, some in the cytoplasm and some in the nucleus. Although individual growth factors may exert profound changes on cells, some effects, such as the induction of cell proliferation, may require the combined actions of more than one growth factor. When the signal pathways are inappropriately activated, cells may lose control of growth and form a tumor. When the pathways are blocked, essential cell types may degenerate and die. It follows that errors in signaling can result in diverse diseases.

In the case of the nervous system, transmembrane signals induced by neurotransmitters regulate the properties of neurons. These small molecules, released by neurons at synapses, bind to receptors on postsynaptic target cells. Neurotransmitter stimulation of target neurons is a critical step in the rapid transmission of nerve impulses, but it can also more slowly regulate the activities of specific genes, allowing the nervous system to modify its properties in response to its environment. Such modification may underlie the processes of neural adaptation and memory.

Our laboratory has identified a group of immediate early-response genes that are rapidly induced by growth factor stimulation and appear to be primary targets in the nucleus for the growth factor–induced signals. Our work focuses on the c-*fos* gene, which is very rapidly induced by a wide range of transmembrane signals. This gene encodes a protein, c-Fos, which is a member of a family of transcription factors that bind to specific sites in the regulatory regions of other genes and thereby control their activity. In this manner c-Fos acts as an intermediary for the conversion of short-term transmembrane signals into longer-term changes in the cell. These studies are supported by the American Cancer Society.

We are particularly concerned with the role of c-*fos* in programs of neuronal differentiation induced by nerve growth factor (NGF). *In vivo*, NGF is required for the differentiation and maintenance of peripheral neurons. Expression of c-*fos* appears to be a first step in the activation of a multistage gene expression program induced by NGF that can culminate in cell division or the induction of terminally differentiated functions.

In our model neuronal cells, we have shown that neurotransmitters as well as growth factors can also induce the expression of c-Fos. It is apparent from this and other studies that c-Fos has a critical role in the adult nervous system, not just in neural development.

We have shown that c-Fos induced by NGF can cooperate with a second protein factor, c-Jun, to induce the gene for tyrosine hydroxylase (TH), an enzyme that catalyzes a critical step in the production of neurotransmitters in the catecholamine family. The combined c-Fos and c-Jun form a heterodimer that binds to a TH gene regulatory

element and induces expression. This participation of c-Fos in TH control may allow a neuron to coordinate the production of catecholamine neurotransmitters with the activity of the neuron during its function in the nervous system. Indeed, stimuli provided by light or smell may use this pathway to control neuronal activity.

The c-Fos protein belongs to a family of proteins including FosB and Fra1 and Fra2, all of which may bind to c-Jun or to members of the Jun protein family. The resulting complexes in turn may all bind to the same DNA element. The different c-Fos family members differ in their patterns of expression following cell stimulation, as well as in their structures outside the DNA-binding domain. This suggests that they may regulate gene activity differentially when complexed with a DNA element such as the TH gene regulatory element.

Recent studies in our laboratory indicate that the TH gene is repressed by a mechanism in which the c-Fos protein, an activator, is replaced by a different c-Fos family member, which serves as a repressor. Indeed, other Fos family members, including FosB, become the predominant species as c-Fos levels dwindle and TH transcription is shut off.

We have gone on to show that NGF induces other genes as well. One of these encodes peripherin, a neuron-specific intermediate filament protein that is present in the axons of peripheral neurons as a component of the neuronal cytoskeleton. Our studies of the developing rat nervous system indicate that peripherin expression coincides with the final steps of neuronal maturation and acquisition of function.

The mechanisms that control peripherin expression appear to be quite distinct from those that control c-Fos or TH. We detect no binding site for the Fos-Jun complex in the peripherin gene. Instead, a negative regulatory element appears to release an inhibitory factor, thus activating the gene.

Study of these mechanisms may give a clue to how cells permanently exit from the cell cycle and induce the expression of genes that they employ after losing the capacity to proliferate. One event that may block exit from the cell cycle is the loss of control of expression of c-*myc*, another growth factor–induced gene. It encodes a protein, c-Myc, that is distantly related to c-Fos and has specialized, but poorly understood, function in inducing cell proliferation. It is expressed at abnormally high levels in many tumors.

We have identified a DNA nucleotide sequence to which c-Myc can bind and a protein partner of c-Myc called Myn that can stimulate its DNA binding. Our studies of c-Myc and Myn and their complexes with DNA indicate they are controlled at many levels, including modification of the mature proteins by phosphorylation. Our studies seek to reveal the role of c-Myc in normal cell proliferation and in tumorigenesis. These studies are supported by the National Institutes of Health.

We find that c-Myc expression is elevated in the naturally occurring childhood brain tumor medulloblastoma. The tumor cells have been blocked to differentiation and have proliferated abnormally. We also find that the transforming gene of adenovirus, E1a, like c-*myc*, can block the differentiation of our model neurons, mimicking the state of the tumor.

We are introducing various sorts of mutants of the c-Myc protein into cells to disrupt pathways stimulated by the normal, wild-type c-Myc. These experiments may reveal other protein factors that cooperate with c-Myc in controlling cell proliferation and differentiation. We seek to understand how cells coordinate their functional maturation with a loss of ability to proliferate as they undergo the final stages of differentiation.

Cell-Cell Interactions Determine Cell Fate in the Drosophila *Retina*

S. Lawrence Zipursky, Ph.D.—Associate Investigator

Dr. Zipursky is also Associate Professor of Biological Chemistry at the University of California School of Medicine, Los Angeles. He received his Ph.D. degree from Albert Einstein College of Medicine, where he studied mechanisms of DNA replication in bacteria in the laboratory of Jerard Hurwitz. He moved to the California Institute of Technology to pursue postdoctoral studies in Drosophila *neurogenetics with Seymour Benzer. He has been on the faculty of UCLA for seven years.*

WE are interested in two questions in developmental biology. First, How are specific cell fates established through cell-cell interactions? And second, What are the mechanisms underlying the specificity of neuronal connectivity? To address these issues, we have focused our studies on the development of the *Drosophila* visual system. I will briefly describe progress we have made in understanding the cellular and molecular mechanisms underlying the specification of a unique cell fate in the developing retina. This work is supported in part by a grant from the National Institutes of Health. We are using similar genetic and molecular approaches toward understanding how different classes of photoreceptor neurons identify their unique postsynaptic targets in the developing brain.

Development of the *Drosophila* Retina

The *Drosophila* retina has a near-crystalline structure of some 800 identical units called ommatidia. Each ommatidium contains a group of cells organized in a stereotyped fashion, with eight photoreceptor neurons, designated R1–R8, forming its core, surrounded by accessory cells that produce screening pigments and the lens. Genetic studies in the 1970s established that there were no strict cell lineage relationships between the cells in the fly's eye. Observations have supported the view that the developmental mechanisms regulating the acquisition of cell identity are dependent not on a cell's ancestry but rather on its interactions with other developing cells.

The R cells are the first to develop within the ommatidial unit. Their stereotyped and sequential pattern of differentiation suggested to early workers that cell fates were established as a consequence of a cascade of inductive interactions. Cell fates were proposed to be a result of unique inductive cues provided by differentiating neighboring cells. The first cell to differentiate is R8, followed by the cells surrounding it. The last cell to differentiate is R7. It contacts the differentiating R8, R1, and R6 cells as well as a number of other unpatterned and undifferentiated cells. It was proposed that the unique R7 cell fate was induced by specific signals requiring direct cell contacts between the R7 precursor and the R1, R6, and R8 cells.

An Inductive Event Specifying R7 Development

The first step toward a molecular description of the developmental mechanisms regulating R7 development was the discovery of the *sevenless* (*sev*) mutation. In flies carrying this mutation, the R7 cell fails to assume its normal fate and becomes a nonneuronal lens-secreting cone cell. The *sev* gene was shown to encode a receptor tyrosine kinase, which is expressed in the R7 precursor cell and many other cells of the developing eye. The molecular structure of the sev protein, its expression pattern, and its genetic requirement in the R7 precursor cell led to the proposal that it is a receptor for an inductive cue specifying an R7 cell fate.

Several years ago we identified a mutation in another gene, *bride of sevenless* (*boss*), which resulted in a phenotype identical to *sev*. Genetic studies indicated that *boss* is required in R8, not for its own development but for that of the R7 cell. This raised the intriguing notion that the boss protein is an inductive ligand to which the sev receptor binds. Molecular analysis of the boss protein revealed that it is an integral membrane protein with a large extracellular domain, multiple transmembrane segments, and a short cytoplasmic tail. Using antibodies, we showed that boss is specifically expressed in the R8 neuron.

Three observations argue that the boss protein is a ligand for the sev receptor. First, mixtures of sev and boss-expressing tissue culture cells bind specifically to one another. Second, membranes containing the boss protein rapidly and specifically activate the sev tyrosine kinase activity. And finally, boss was shown to be transferred from the surface of the R8 cell to an organelle in R7 referred to as a multivesicular body. This internalization is strictly dependent upon the presence of the sev protein on the surface of the R7 cell.

Earlier workers had demonstrated that the sev

protein is expressed on many cells within an ommatidium in addition to the R7 precursor cell. Nevertheless, only one cell assumes an R7 cell fate. Four of these additional cells contact the boss-expressing R8 cell, and the others are one or more cell diameters away. We have shown that multiple mechanisms limit the induction to only one cell, through studies in which we examined R7 development in various mutant backgrounds and in response to ectopically expressed boss protein. These studies revealed that the cells that do not contact the R8 cells are competent to respond to the inductive cue and do so if presented with a surface-bound form of boss. Conversely, the sev-expressing cells that normally contact R8 are restricted from responding to the cue as a result of a very early developmental event that commits them to alternative fates.

We believe that similar molecular and developmental strategies involving the carefully controlled expression of surface-bound ligands and competing alternative developmental pathways will be important in establishing the remarkable cellular diversity and organization of more-complex nervous systems, such as the vertebrate brain.

Molecular Genetics of Sensory Transduction

Charles S. Zuker, Ph.D.—Associate Investigator

Dr. Zuker is also Associate Professor of Biology and of Neurosciences at the University of California School of Medicine, San Diego. He received his Ph.D. degree from the Massachusetts Institute of Technology for studies with Harvey Lodish. He carried out postdoctoral research with Gerald Rubin in the Department of Biochemistry at the University of California, Berkeley, before joining the Department of Biology at UCSD. Dr. Zuker is currently a Pew Fellow in the Biomedical Sciences and a Fellow of the March of Dimes Foundation.

AN understanding of signal transduction is essential to elucidating the cellular and molecular basis of information processing in biological systems. The primary event in the processing of visual stimuli is phototransduction, the conversion of light energy into a change in the ionic permeabilities of the photoreceptor cell membrane. The aim of our research is to clarify mechanisms used for signal transduction in the visual system, using a combined molecular, genetic, and physiological approach. The study of this process in *Drosophila* applies powerful molecular genetic techniques to identify novel transduction molecules and to examine their function *in vivo*, in their normal cellular and organismal environment.

Experimental Strategy

Over the past few years my colleagues and I have been working on the isolation and characterization of genes important for photoreceptor cell function. Most recently we have focused on genes encoding proteins involved in the regulation of the visual transduction cascade. We have identified a protein kinase C (PKC) that is expressed exclusively in the *Drosophila* visual system. Analysis of the light response from mutants defective in this PKC showed it to be required for the deactivation and rapid desensitization of the visual cascade. The availability of a PKC mutant in *Drosophila* provides the basis for genetic and biochemical studies to identify biologically relevant substrates and regulators of this enzyme.

Rhodopsin, like other G protein–coupled receptors, is phosphorylated by a specific kinase upon activation (rhodopsin kinase). This reaction is thought to be involved in the termination of rhodopsin's active state. Phosphorylation, however, is not sufficient for receptor inactivation, since the phosphorylated form can still activate the G protein. Complete inactivation has been shown (*in vitro*) to require the stoichiometric interaction of rhodopsin with a protein called arrestin, or S antigen. This protein has also been implicated in a number of autoimmune retinal disorders in mammals. We have isolated the genes encoding two photoreceptor cell–specific arrestin molecules in *Drosophila* and have generated mutants of these genes. Our results should help us understand the molecular basis of G protein–coupled receptor regulation and of relevant abnormalities in the human visual and nervous systems.

Mechanotransduction

We have recently begun to study mechanotransduction, the process by which specialized sensory cells convert mechanical stimuli—for instance, sound, touch, gravity, or movement—into electrical (neuronal) signals. In contrast to phototransduction, nothing is known about the molecular basis of mechanosensitivity. Our aim is to identify genes and proteins involved in mechanotransduction by isolating mutations that affect mechanosensory behavior.

When *Drosophila* larvae are touched gently, they contract and retreat. We have developed a screen for mutant larvae defective in this behavior and have isolated several mutant lines. Mutations that specifically affect the working of the sensory organs are of greatest interest. Mapping the mutations—in prelude to isolating and cloning the affected genes—is in progress. In addition, we have developed a molecular approach to isolate genes expressed specifically in mechanosensory organs. We are using developmental mutants that overproduce or lack mechanosensory bristles to isolate such genes by subtractive hybridization.

Investigators by Location

ALABAMA

Birmingham
University of Alabama at Birmingham and associated hospitals

Cooper, Max D., M.D.

CALIFORNIA

Berkeley
University of California, Berkeley, and associated hospitals

Goodman, Corey S., Ph.D.
Rubin, Gerald M., Ph.D.
Schekman, Randy W., Ph.D.
Tjian, Robert, Ph.D.

Los Angeles
University of California, Los Angeles, and associated hospitals

Kaback, H. Ronald, M.D.
Simpson, Larry, Ph.D.
Smale, Stephen T., Ph.D.
Witte, Owen N., M.D.
Zipursky, S. Lawrence, Ph.D.

University of Southern California and associated hospitals

Lai, Michael M.-C., M.D., Ph.D.

Palo Alto
Stanford University and the Stanford University Hospital

Aldrich, Richard W., Ph.D.
Barsh, Gregory S., M.D., Ph.D.
Brown, Patrick O., M.D., Ph.D.
Crabtree, Gerald R., M.D.
Davis, Mark M., Ph.D.
Francke, Uta, M.D.
Goodnow, Christopher C., B.V.Sc., Ph.D.
Kobilka, Brian K., M.D.
Nusse, Roel, Ph.D.
Scheller, Richard H., Ph.D.
Schoolnik, Gary K., M.D.
Weissman, Irving L., M.D.

Pasadena
California Institute of Technology and associated hospitals

Anderson, David J., Ph.D.
Bjorkman, Pamela J., Ph.D.
Sternberg, Paul W., Ph.D.

San Diego
The Salk Institute for Biological Studies

Evans, Ronald M., Ph.D.
Sejnowski, Terrence J., Ph.D.
Stevens, Charles F., M.D., Ph.D.

University of California, San Diego, and the UCSD Medical Center

Bevilacqua, Michael P., M.D., Ph.D.
Emr, Scott D., Ph.D.
Rosenfeld, Michael G., M.D.
Tsien, Roger Y., Ph.D.
Zuker, Charles S., Ph.D.

San Francisco
University of California, San Francisco, and associated hospitals

Agard, David A., Ph.D.
Ganem, Donald E., M.D.
Gitschier, Jane M., Ph.D.
Grosschedl, Rudolf, Ph.D.
Jan, Lily Y., Ph.D.
Jan, Yuh Nung, Ph.D.
Kan, Yuet Wai, M.D., D.Sc.
Littman, Dan R., M.D., Ph.D.
Payan, Donald G., M.D.
Peterlin, B. Matija, M.D.
Reichardt, Louis F., Ph.D.
Sedat, John W., Ph.D.
Weiss, Arthur, M.D., Ph.D.
Williams, Lewis T., M.D., Ph.D.

COLORADO

Boulder
University of Colorado at Boulder and the University's Health Sciences Center

Cech, Thomas R., Ph.D.
Kirkegaard, Karla A., Ph.D.

Denver
National Jewish Center for Immunology and Respiratory Medicine

Kappler, John W., Ph.D.
Marrack, Philippa, Ph.D.

University of Colorado Health Sciences Center and associated hospitals

Maller, James L., Ph.D.

CONNECTICUT

New Haven
Yale University and associated hospitals

Artavanis-Tsakonas, Spyridon, Ph.D.
Bottomly, H. Kim, Ph.D.
Brünger, Axel T., Ph.D.
Cresswell, Peter, Ph.D.

De Camilli, Pietro, M.D.
Flavell, Richard A., Ph.D.
Fox, Robert O., Ph.D.
Ghosh, Sankar, Ph.D.
Horwich, Arthur L., M.D.
Jahn, Reinhard, Ph.D.
Janeway, Charles A., Jr., M.D.
Lerner, Michael R., M.D., Ph.D.
Reeders, Stephen T., M.D.
Schatz, David G., Ph.D.
Sigler, Paul B., M.D., Ph.D.
Steitz, Joan A., Ph.D.
Steitz, Thomas A., Ph.D.
Tanabe, Tsutomu, Ph.D.
Wolin, Sandra L., M.D., Ph.D.

GEORGIA

Atlanta
Emory University School of Medicine and associated hospitals

Warren, Stephen T., Ph.D.

ILLINOIS

Chicago
The University of Chicago and The University of Chicago Hospitals

Bell, Graeme I., Ph.D.
Fuchs, Elaine, Ph.D.
Laimins, Laimonis A., Ph.D.
Lindquist, Susan L., Ph.D.
Singh, Harinder, Ph.D.
Steiner, Donald F., M.D.
Sukhatme, Vikas P., M.D., Ph.D.

Evanston
Northwestern University and associated hospitals

Lamb, Robert A., Ph.D.

INDIANA

Bloomington
Indiana University and associated hospitals

Kaufman, Thomas C., Ph.D.

Indianapolis
Indiana University School of Medicine and associated hospitals

Williams, David A., M.D.

IOWA

Iowa City
University of Iowa and associated hospitals

Campbell, Kevin P., Ph.D.
Donelson, John E., Ph.D.
Welsh, Michael J., M.D.

MARYLAND

Baltimore
The Carnegie Institution of Washington and The Johns Hopkins Hospital

McKnight, Steven Lanier, Ph.D.
Spradling, Allan C., Ph.D.

The Johns Hopkins University and Hospital

Beachy, Philip A., Ph.D.
Corden, Jeffry L., Ph.D.
Craig, Nancy L., Ph.D.
Desiderio, Stephen V., M.D., Ph.D.
Huganir, Richard L., Ph.D.
Nathans, Daniel, M.D.
Nathans, Jeremy, M.D., Ph.D.
Reed, Randall R., Ph.D.
Valle, David L., M.D.
Yau, King-Wai, Ph.D.
Yellen, Gary, Ph.D.

MASSACHUSETTS

Boston
Brigham and Women's Hospital

Chin, William W., M.D.
Cunningham, James M., M.D.
Maas, Richard L., M.D., Ph.D.

The Children's Hospital

Alt, Frederick W., Ph.D.
Kunkel, Louis M., Ph.D.
Nadal-Ginard, Bernardo, M.D., Ph.D.
Orkin, Stuart H., M.D.

Harvard Medical School and Brigham and Women's Hospital

Church, George M., Ph.D.
Duyk, Geoffrey M., M.D., Ph.D.
Leder, Philip, M.D.
Perrimon, Norbert, Ph.D.
Seidman, Jonathan G., Ph.D.

Massachusetts General Hospital

Alexander-Bridges, Maria C., M.D., Ph.D.
Corey, David P., Ph.D.
Habener, Joel F., M.D.

Tufts University School of Medicine and associated hospitals

Isberg, Ralph R., Ph.D.

Cambridge
Harvard College, Arts and Sciences, and The Children's Hospital

Harrison, Stephen C., Ph.D.
Wiley, Don C., Ph.D.

Massachusetts Institute of Technology and associated hospitals

Horvitz, H. Robert, Ph.D.
Hynes, Richard O., Ph.D.
Kim, Peter S., Ph.D.
Lehmann, Ruth, Ph.D.
Pabo, Carl O., Ph.D.
Page, David C., M.D.
Steller, Hermann, Ph.D.
Tonegawa, Susumu, Ph.D.

Waltham
Brandeis University and associated hospitals

Miller, Christopher, Ph.D.
Rosbash, Michael, Ph.D.

Worcester
University of Massachusetts and associated hospitals

Davis, Roger J., Ph.D.

MICHIGAN

Ann Arbor
University of Michigan and associated hospitals

Bonadio, Jeffrey F., M.D.
Collins, Francis S., M.D., Ph.D.
Feinberg, Andrew P., M.D., M.P.H.
Ginsburg, David, M.D.
Kurnit, David M., M.D., Ph.D.
Leiden, Jeffrey M., M.D., Ph.D.
Lowe, John B., M.D.
Nabel, Gary J., M.D., Ph.D.
Thompson, Craig B., M.D.
Wilson, James M., M.D., Ph.D.

MISSOURI

St. Louis
Washington University and associated hospitals

Atkinson, John P., M.D.
Chaplin, David D., M.D., Ph.D.
Holers, V. Michael, M.D.
Korsmeyer, Stanley J., M.D.
Loh, Dennis Y.-D., M.D.
Olson, Maynard V., Ph.D.
Pike, Linda J., Ph.D.
Sadler, J. Evan, M.D., Ph.D.
Thomas, Matthew L., Ph.D.

NEW JERSEY

Princeton
Princeton University and associated medical centers

Shenk, Thomas E., Ph.D.
Tilghman, Shirley M., Ph.D.

NEW YORK

Bronx
Albert Einstein College of Medicine of Yeshiva University and associated hospitals

Bloom, Barry R., Ph.D.
Jacobs, William R., Jr., Ph.D.

Cold Spring Harbor
Cold Spring Harbor Laboratory and associated hospitals

Beach, David H., Ph.D.

New York
Columbia University and associated hospitals

Axel, Richard, M.D.
Hendrickson, Wayne A., Ph.D.
Jessell, Thomas M., Ph.D.
Kandel, Eric R., M.D.
Siegelbaum, Steven A., Ph.D.
Struhl, Gary, Ph.D.

Cornell University Medical College

O'Donnell, Michael E., Ph.D.

Memorial Sloan-Kettering Cancer Center

Massagué, Joan, Ph.D.

New York University (Medical Center and Washington Square) and associated hospitals

Movshon, J. Anthony, Ph.D.
Ziff, Edward B., Ph.D.

The Rockefeller University and Rockefeller University Hospital

Blobel, Günter, M.D., Ph.D.
Burley, Stephen K., Ph.D.

Desplan, Claude, Ph.D.
Friedman, Jeffrey M., M.D., Ph.D.
Geliebter, Jan, Ph.D.
Heintz, Nathaniel, Ph.D.
Kuriyan, John, Ph.D.
Nussenzweig, Michel C., M.D., Ph.D.
Sakmar, Thomas P., M.D.
Young, Michael W., Ph.D.

Stony Brook
State University of New York at Stony Brook and University Hospital at Stony Brook

Adams, Paul R., Ph.D.

NORTH CAROLINA

Durham
Duke University, including Duke University Medical Center

Bennett, G. Vann, M.D., Ph.D.
Blackshear, Perry J., M.D., D.Phil.
Caron, Marc G., Ph.D.
Cullen, Bryan R., Ph.D.
Davis, Laura I., Ph.D.
Lefkowitz, Robert J., M.D.
Nevins, Joseph R., Ph.D.
Parker, Keith L., M.D., Ph.D.

OKLAHOMA

Oklahoma City
Oklahoma Medical Research Foundation and associated hospitals

Esmon, Charles T., Ph.D.

OREGON

Eugene
University of Oregon and associated hospitals

Matthews, Brian W., Ph.D.

PENNSYLVANIA

Philadelphia
University of Pennsylvania and associated hospitals

Brugge, Joan S., Ph.D.
Dreyfuss, Gideon, Ph.D.
Kadesch, Thomas R., Ph.D.
Liebhaber, Stephen A., M.D.
Malim, Michael H., Ph.D.
Nussbaum, Robert L., M.D.
Taub, Rebecca A., M.D.

TENNESSEE

Memphis
St. Jude Children's Research Hospital

Sherr, Charles J., M.D., Ph.D.

Nashville
Vanderbilt University, including Vanderbilt University Hospital

Exton, John H., M.D., Ph.D.

TEXAS

Dallas
University of Texas Southwestern Medical Center at Dallas and associated hospitals

Beutler, Bruce A., M.D.
Deisenhofer, Johann, Ph.D.
Fischer Lindahl, Kirsten, Ph.D.
Garbers, David L., Ph.D.
Gething, Mary-Jane H., Ph.D.
Sprang, Stephen R., Ph.D.
Südhof, Thomas C., M.D.
Yanagisawa, Masashi, M.D., Ph.D.

Houston
Baylor College of Medicine and associated hospitals

Beaudet, Arthur L., M.D.
Bellen, Hugo J., D.V.M., Ph.D.
Belmont, John W., M.D., Ph.D.
Caskey, C. Thomas, M.D.
Cohen, Stephen M., Ph.D.
Ledley, Fred D., M.D.
Overbeek, Paul A., Ph.D.
Quiocho, Florante A., Ph.D.
Soriano, Philippe M., Ph.D., D.Sc.
Woo, Savio L. C., Ph.D.

Rice University and associated hospitals

Gomer, Richard H., Ph.D.

UTAH

Salt Lake City
University of Utah, including University of Utah Medical Center

Capecchi, Mario R., Ph.D.
Gesteland, Raymond F., Ph.D.
Lalouel, Jean-Marc, M.D., D.Sc.
Sakonju, Shigeru, Ph.D.
Thummel, Carl S., Ph.D.
White, Raymond L., Ph.D.

WASHINGTON

Seattle

Fred Hutchinson Cancer Research Center

Henikoff, Steven, Ph.D.
Weintraub, Harold M., M.D., Ph.D.

University of Washington
and associated hospitals

Bevan, Michael J., Ph.D.
Glomset, John A., M.D.
Hurley, James B., Ph.D.
Palmiter, Richard D., Ph.D.
Perlmutter, Roger M., M.D., Ph.D.

WISCONSIN

Madison

University of Wisconsin
and associated hospitals

Carroll, Sean B., Ph.D.

International Research Scholars

In recognition of the contributions of scientists outside of the United States to advances in biomedical science, the Institute initiated in 1991 its International Research Scholars Program. This is a limited experimental effort that provides research support through five-year grants to promising scientists working in areas of fundamental biomedical research related to the Institute's ongoing medical research programs.

The investigators invited to compete for the initial grant awards were located in Canada and Mexico. From those invited to compete, 24 were selected (14 in Canada, 10 in Mexico). Grants also were awarded to the Mexican Academia de la Investigacion Cientifica and to the United States National Academy of Sciences for joint activities over four years to promote the exchange of scientific information and encourage cooperation between the scientific communities in each country, particularly in the life sciences.

The next phase of the international program will focus on biomedical scientists in the United Kingdom, Australia, and New Zealand. The emphasis remains on the support of outstanding scientists whose research careers are still developing, rather than those in the later phases of a distinguished career. Grant awards for this second group will be announced late in 1992.

Canada

Bernstein, Alan, Ph.D.
Finlay, B. Brett, Ph.D.
Greenblatt, Jack, Ph.D.
Grinstein, Sergio, Ph.D.
Gros, Philippe, Ph.D.
Joyner, Alexandra L., Ph.D.
McGhee, James D., Ph.D.
Mosmann, Tim R.., Ph.D.
Pawson, Tony, Ph.D.
Read, Randy J., Ph.D.
Rossant, Janet, Ph.D.
Roy, Jean-Pierre, M.D.
Snutch, Terry P., Ph.D.
Tsui, Lap-Chee, Ph.D.

Mexico

Arias, Carlos F., Ph.D.
Calva, Edmundo, Ph.D.
Cota, Gabriel, Ph.D.
Darszon, Alberto, Ph.D.
Guarneros Peña, Gabriel, Ph.D.
Herrera-Estrella, Luis R., Ph.D.
Lizardi, Paul M., Ph.D.
Orozco, Esther, Ph.D.
Possani, Lourival Domingos, Ph.D.
Romo, Ranulfo, M.D., Ph.D.

Molecular Biology and Epidemiology for Control of Rotavirus Diarrhea

Carlos F. Arias, Ph.D.—International Research Scholar

Dr. Arias is Investigador Titular B, Department of Molecular Biology, at the Biotechnology Institute, National Autonomous University of Mexico, Cuernavaca. He received his undergraduate degree in biochemistry and his M.S. and Ph.D. degrees in biomedical research from the National University of Mexico, Mexico City. He recently conducted sabbatical research with James Strauss at the California Institute of Technology.

ACUTE, infectious diarrhea is the commonest cause of morbidity and mortality among young children living in developing countries, accounting for as many as one billion illnesses and between four and five million deaths annually. Rotaviruses are the leading cause of severe diarrheal disease in children under three years of age, and it is estimated that an effective vaccine would save about 800,000 children's lives per year.

Because rotavirus plays such an important role in severe infantile gastroenteritis, and because even advanced levels of hygiene seem unable to control the spread of rotavirus infections, there has been considerable interest in developing effective vaccination strategies.

Our laboratory is interested in studying how rotaviruses attach and enter their host cell and how they replicate there to produce the viral progeny. Also among our interests are the host immune response to rotavirus infection and the epidemiology of these viruses in Mexico.

As in other infectious agents, proteins located at the surface of the rotavirus particles are involved in the early interactions (attachment and penetration) of the virions with the host cell. These proteins contain antigenic determinants that represent major immunological targets. The structural and functional characterization of the viral surface proteins, as well as the study of immune response in the infected host, should contribute to the development of rationally designed therapeutic agents and new prevention strategies. Moreover, the success of a vaccine may depend on knowledge of the epidemiology of the pathogen in the particular geographic area where the vaccine is to be used.

The surface of the rotavirus contains two proteins, VP4 and VP7. VP4 forms spikes that extend from the viral surface. This protein has been associated with a variety of viral functions, including the determination of virulence *in vivo* and the ability to agglutinate red blood cells (hemagglutination). VP4 is also important in the penetration of the virion into the cell. On the other hand, the protein responsible for the initial attachment of rotavirus to its host cell remains controversial; both VP4 and VP7 have been proposed.

The attachment of animal rotaviruses to cells in culture is mediated by compounds containing sialic acid, since treatment of cells with sialidases greatly reduces the binding of virus particles to the cell surface. In addition, the hemagglutinating activity of rotaviruses and their binding to cultured epithelial cells can also be inhibited by incubation of the virus with sialoglycoproteins such as glycophorin.

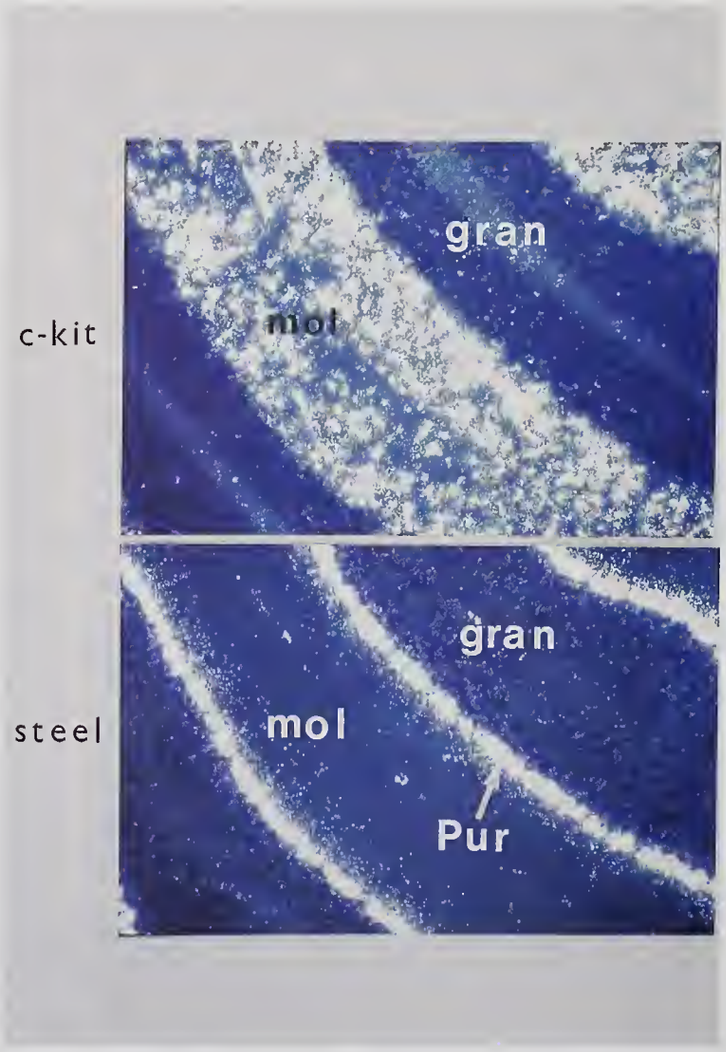
We have isolated rotavirus mutants whose binding is no longer inhibited by treatment of cells with sialidases or incubation with glycophorin. The detailed analysis of these mutants should give us information about the protein(s) involved in the attachment of the virions to the cell. Furthermore, we are constructing deletion mutants and chimeras between VP4 genes of hemagglutinating and nonhemagglutinating rotaviruses, to define further the protein domains involved in the binding of the virus to epithelial and red blood cells.

After the virus's initial attachment to the cell surface, the next step in its infection cycle is to enter the cell. The entry of the virus particle can be augmented by treating it with trypsin, and probably is trypsin dependent. Experiments in progress have shown that this proteolytic treatment induces three specific cleavages of VP4. We are interested in identifying the cleavage(s) directly associated with enhancement of viral infectivity and in learning about the mechanism through which the cleaved VP4 protein mediates the virion's penetration of the cell.

Both surface proteins have also been characterized immunologically. It has been found that antibodies to either protein neutralize the virus *in vitro* and are capable of passively protecting animals from rotavirus challenge. In addition, oral infection with live rotavirus stimulates protective immunity, which can be mediated by VP4 and/or VP7. These observations make the two proteins attractive candidates for recombinant subunit vaccines.

Rotavirus infections are limited to the gut, and it is believed that a vaccine will have to be given orally to stimulate a local (mucosal) immune response in order to induce protection. An attractive approach to this, alternative to the use of live attenuated virus, is the use of live oral vaccines containing nonpathogenic strains of enteric bacteria carrying the genes that code for the protective rotavirus antigens. We are currently constructing expression vectors to direct the synthesis of the surface proteins of rotavirus in the enteric bacteria *Salmonella* and *Lactobacillus*, and will test the potential of these recombinant strains to induce a protective immune response against rotavirus infection.

Finally, our laboratory is also interested in the study of some aspects of the epidemiology of rotavirus. Six different serotypes of human rotaviruses have been identified, four of which appear to account for the majority of isolates. We have found that the four major serotypes circulate in Mexico. We are currently studying the molecular determinants of virulence of rotavirus and investigating how the frequency of the different serotypes changes over time and in relation to virulence.



In situ *localization of* c-kit *and* Steel *expression in adult mouse cerebellum. The protein products of the proto-oncogene* c-kit *and its ligand, the protein product of the* Steel *gene, interact to activate Kit tyrosine kinase activity vital to the function of melanocytes, blood cells, and germ cells. The abbreviations identify the molecular layer (mol), granular layer (gran), and Purkinje cells (Pur).*

From Motro, B., van der Kooy, D., Rossant, J., Reith, A., and Bernstein, A. 1991. Development *113:1207–1221.*

Molecular Genetics of Normal and Leukemic Hematopoiesis

Alan Bernstein, Ph.D.—International Research Scholar

Dr. Bernstein is Head of the Division of Molecular and Developmental Biology and Associate Director of the Samuel Lunenfeld Research Institute at Mount Sinai Hospital, Toronto. He received his Ph.D. degree in medical biophysics from the University of Toronto and trained as a postdoctoral fellow at the Imperial Cancer Research Fund Laboratories, London. Before moving to Mount Sinai he was Senior Scientist in the Division of Biological Research at the Ontario Cancer Institute. He is the first incumbent of the Anne Tanenbaum Chair in Molecular and Developmental Biology and was recently elected a Fellow of the Royal Society of Canada.

MY laboratory is interested in the molecular mechanisms that govern the orderly production of hematopoietic cells in the adult mammal. It is clear from earlier work involving the generation of unique clonal markers (either visible cytogenetic markers or the chromosomal integration sites of retroviral vectors) that hematopoiesis involves the proliferation and differentiation of pluripotent stem cells. These cells have the developmental and proliferative capacity to produce the millions of myeloid and lymphoid cells that are continuously required in adult life. Thus the hematopoietic system has served as an attractive and important experimental model of developmental processes. Defects in hematopoiesis can also lead to disease, including anemias, leukemias, and such genetic disorders as sickle cell anemia and thalassemia, making the understanding of this system important to both medicine and biology.

We are taking two main approaches to elucidate the mechanisms that control the regulation of hematopoietic stem cells. The first approach involves the analysis of mouse mutations that disrupt hematopoiesis, while the second involves the study of the multistage erythroleukemias induced by the various forms of Friend murine leukemia virus, a murine retrovirus.

The *white-spotting* (*W*) and *Steel* (*Sl*) Loci

Mutations at either the dominant *white-spotting* (*W*) or *Steel* (*Sl*) loci can lead to coat color defects, severe macrocytic anemia, and sterility. This similarity in pleiotropic phenotype of *W* and *Sl* mutants is striking, particularly as there is no common developmental origin of cells that give rise to melanocytes, blood cells, and germ cells. Furthermore, the *W* and *Sl* loci are clearly distinct genes, as they map on mouse chromosomes 5 and 10, respectively.

Initial insights into their mechanism of action came from *in vivo* cell-mixing experiments, involving either bone marrow transplantation or aggregation chimeras with animals of different genotypes. These early studies established that the developmental defects in *W* mutant mice result from a cell-autonomous, intrinsic stem cell defect, whereas the cellular defect in *Sl* mutant mice is in the microenvironment in which these cells develop in the embryo and function in the adult animal.

Several years ago, we showed that *W* was allelic with the proto-oncogene c-*kit*, a member of the family of genes that includes c-*fms* and *pdgfr*, which encode transmembrane receptors with tyrosine kinase activity. Different *W* alleles are either deletions or point mutations in the cytoplasmic tyrosine kinase domain that result in partial or complete loss of tyrosine kinase activity and a consequent reduction in the signaling capability of the Kit receptor. Thus the *W* mutant mouse provided the first example of a germline mutation in a mammalian proto-oncogene. This work is supported by a grant from the National Institutes of Health.

These findings predicted that *Sl* might encode the ligand for the Kit receptor and that this ligand might be a potent growth or survival factor for melanoblasts, germ cells, and hematopoietic stem cells. The microenvironmental nature of the defects in *Sl* mutant mice also predicted that the *Sl* protein product might be membrane bound. All three predictions have turned out to be correct. Thus the *W-Sl* gene pair has provided strong molecular genetic evidence for the important role that cell-cell interactions play in the development and function of hematopoietic and other stem cells. Their protein products are strikingly similar in design and function to those of the *sev* and *boss* genes, which control the development of the R7 photoreceptor cell in the compound eye of the fruit fly *Drosophila*.

Friend Leukemia Virus

In the second approach, we are identifying and analyzing cellular genes important in the induction of the multistage erythroleukemias induced by Friend leukemia virus. Several years ago, we showed that the *p53* gene was inactivated by deletion or retroviral insertion in approximately 30

percent of independent leukemic clones isolated from the spleens of mice infected with Friend virus. These experiments suggested that *p53* is a tumor-suppressor gene, a conclusion supported by recent experiments demonstrating allelic loss and mutation in the *p53* gene in a broad spectrum and high proportion of human cancers.

The importance of *p53* in the evolution of Friend leukemia also became evident from studies on transgenic mice that express high levels of mutant forms of the *p53* protein. In addition to displaying an increased spontaneous predisposition to a variety of malignancies, these *p53* transgenic mice are more susceptible to the late stages of Friend leukemia. Thus inactivation of the *p53* gene, either somatically after Friend virus infection or as the result of inheritance of dominant negative alleles of *p53*, appears to be a central event in the disease. These studies also suggested that it is the accumulation of a specific set of mutational events, rather than the order in which they normally occur, that is critical for leukemia induction.

In addition to inactivation of the *p53* gene, activation of one of two novel members of the *ets* gene family of DNA-binding transcriptional activators occurs during the evolution of Friend leukemia. Transcription of the *Spi*-1 gene is activated as a result of the integration of Friend spleen focus-forming virus (SFFV) in 95 percent of erythroleukemic clones induced in adult mice by SFFV. In contrast, we showed last year that transcription of another *ets* family member, *Fli*-1, is activated by insertion of the replication-competent Friend MuLV in 75 percent of erythroleukemia clones induced after infection of newborn mice. The members of the *ets* gene family, which include c-*ets-1* and c-*ets-2*, *elk-1* and *elf-1*, *erg*, *E74*, *Spi*-1/PU.1, and *Fli*-1, contain a conserved domain of 80–90 amino acids, the ETS domain, which is involved in specific DNA binding, as well as a less-conserved transcriptional activation domain.

The strict specificity of the integration sites for SFFV and Friend MuLV is intriguing, as *Spi*-1 and *Fli*-1 are both members of the same gene family and both SFFV and Friend MuLV induce erythroleukemias involving activation of the receptor for erythropoietin and inactivation of the *p53* gene. We are currently pursuing the hypothesis that *Fli*-1 and *Spi*-1 are functionally distinct genes, encoding proteins that transactivate a distinct set of genes downstream in the leukemogenic pathway. These studies involve the analysis of the DNA-binding specificity of the *Fli*-1 and *Spi*-1 proteins as well as the generation of mutant mice carrying either gain- or loss-of-function mutations in these genes. This work is supported by a grant from the National Cancer Institute of Canada.

The *ets* gene family was first discovered by analyzing the genome of the avian erythroblastosis virus E26, which contains a part of a *myb* fusion protein, the v-*ets* oncogene. Thus at least three members of the *ets* family, v-*ets*, *Spi*-1, and *Fli*-1, are involved in the leukemic transformation of erythroid cells. This specificity of *ets* genes for the erythroid lineage suggests that members of this gene family are involved in the regulation of one or more genes critical to erythropoiesis. We are interested in identifying these genes to gain further insights into both the regulation of normal hematopoiesis and the induction of leukemia.

Molecular Biology of Two Enteropathogenic Bacteria

Edmundo Calva, Ph.D.—International Research Scholar
Dr. Calva is Associate Professor and Chairman of Molecular Biology at the Biotechnology Institute, National Autonomous University of Mexico (UNAM), Cuernavaca. He received his Ph.D. degree in molecular biology from the University of Wisconsin–Madison under Richard Burgess and did postdoctoral research at UNAM. He is currently President of the Mexican Biochemical Society.

THE study of infectious diseases provides opportunities to explore various basic biological phenomena. During an infection, enteropathogenic bacteria participate in numerous biological events that may lead to intestinal illness. Upon oral ingestion, the bacteria must undergo a series of interactions with the host that involve specific molecules, either on the surface or in the interior of the bacterial and host cells. For instance, bacteria adhere to epithelial gut cells and produce enterotoxins that result in diarrhea. They can invade host tissue with the same effect. Invasion sometimes results in a systemic infection, in which the bacteria have means of protection against the immune system, seriously jeopardizing health.

Current research questions in the area of bacterial pathogenesis address the molecular mechanisms of the bacteria-host interaction. Our knowledge of adherence, invasion, resistance to the immune system, and enterotoxin production, among other processes, is just being unveiled. We know only scant details of some of these processes in a few bacteria, and even then are not completely aware of the genetic variation within the species studied—variation that may result in different strains with varying capabilities for causing disease. Furthermore, some bacterial and parasite antigens share structures with stress proteins, like those expressed during heat shock. An open area of study involves defining the global genetic circuits that regulate virulence factors and determining whether these "regulons" share common features with those encountered in the response to stresses other than infection.

Health biotechnology should benefit from the definition of bacterial antigens that have a role in infection or against which a specific immune response is mounted, thus permitting the development of better vaccines and diagnostic procedures as well as furthering our understanding of the structure and function of the immune system. Benefits should also emerge from the isolation and characterization of specific bacterial genes, together with knowledge about their distribution and polymorphism in bacterial populations. Such information should permit the rapid and specific detection of bacteria through nucleic acid amplification procedures. These can not only be useful for monitoring infections in animals, plants, or humans, or the contamination of foodstuff, but should also be valuable in the definition of modes of transmission and environmental reservoirs for the bacteria. In this manner, molecular epidemiology will very likely shed light on yet other biological phenomena.

Salmonella typhi

S. typhi is the causal agent of typhoid fever (TF) in humans, a disease estimated to afflict annually more than 12 million persons worldwide. TF is the result of a systemic infection, in which *S. typhi* can be isolated from blood cultures. As a gram-negative bacterium, *S. typhi* has an outer membrane that surrounds an inner one and the cell wall. Thus exposed to the cell's exterior environment, outer membrane proteins (OMPs) have been shown to be important immunogens for protection against various bacterial infections in laboratory models. We have demonstrated the utility of OMPs for the rapid immunodiagnosis of TF in patients from different parts of the world, most of whom raise specific antibodies to OMP preparations.

Our laboratory has reported the isolation and characterization of an *S. typhi* gene coding for OmpC, a major OMP. Amino acid sequence alignment with other OMPs that form pores (porins) has allowed structure prediction in the porin superfamily. Using site-directed mutagenesis, we have constructed a gene that codes for a chimeric protein containing a foreign epitope inserted in a region predicted to be exposed on the cell surface. This epitope is a segment of a capsid protein from rotavirus, a causal agent of diarrhea, against which neutralizing antibodies are made. The chimeric protein does indeed contain this epitope on a segment exposed at the bacterial cell surface, supporting the notion that the native region is located toward the exterior and indicating that OmpC could be used as a carrier of heterologous epitopes. This information on protein topology

can be useful in the design of multivalent vaccines against several infectious agents.

We have used the *S. typhi ompC* gene to probe the genetic variability of the salmonella genus. In this manner we have been able to establish some phylogenetic relationships among several species.

We have also determined that expression of *S. typhi* OmpC is influenced differently by medium osmolarity than its well-studied counterpart in *Escherichia coli*. Nevertheless, expression of *S. typhi ompC* is also dependent on the *E. coli* OmpR transcriptional activator. Apparently the two bacteria differ in the mechanisms of gene expression in response to osmotic stress, although common effectors appear to be shared. We are presently studying this phenomenon in detail.

In addition, we have isolated other *S. typhi* OMP genes, namely *ompF* and *phoE*, and are characterizing these to gain a better understanding of structure-function relations in porins and their genes.

Campylobacter jejuni

C. jejuni is one of the major causal agents of diarrhea throughout the world. Molecular biology of this organism has developed slowly, mostly because of the difficulty of maintaining stably cloned *C. jejuni* DNA segments in *E. coli*. Consequently, only a handful of *C. jejuni* genes have been isolated. The mechanisms underlying this DNA instability are not understood, and research in this area might well yield valuable results.

The isolation of *C. jejuni*, producing a cholera-like enterotoxin, has been associated with clinical symptoms of a watery secretory type of diarrhea. We have shown that the *C. jejuni* chromosome contains sequences similar to the enterotoxin genes of *Vibrio cholerae* (CT enterotoxin) and *E. coli* (LT, heat-labile enterotoxin). In cloning and characterizing these sequences, we have explored different host cells and different-size *C. jejuni* DNA fragments in order to find a successful cloning procedure. In the process, we have isolated two cryptic fragments, one highly specific for *C. jejuni* and *C. coli* (another diarrhea-producing campylobacter) and the other revealing genetic variability between the two bacteria. We are testing the usefulness of both DNA probes in field epidemiology and characterizing them at the nucleotide level.

Functional Heterogeneity in Prolactin-secreting Cells

Gabriel Cota, Ph.D.—International Research Scholar

Dr. Cota is Professor of Physiology, Biophysics, and Neurosciences at the Center for Research and Advanced Studies, National Polytechnic Institute, Mexico City. He received his Ph.D. degree in physiology and biophysics from the Center and subsequently carried out postdoctoral research with Clay M. Armstrong at the University of Pennsylvania, as a Fogarty International Research Fellow.

OUR laboratory is interested in the cellular mechanisms involved in the control of prolactin secretion. Prolactin is a vertebrate hormone that participates in the regulation of a diversity of physiological processes, including lactation. This versatile chemical messenger is produced in the pituitary gland by endocrine cells called lactotropes or mammotropes. Basic information about lactotrope function and its control offers insight into the pathogenesis of hyperprolactinemia, a frequent hypothalamic pituitary disorder in humans.

Until recently, lactotropes were commonly thought to comprise a homogeneous cell population in the normal pituitary gland. However, studies performed on cultured pituitary cells indicate the existence of a considerable lactotrope heterogeneity. In these accessible model systems, subsets of lactotropes that differ in basal secretory activity or responsiveness to extracellular regulatory factors have been distinguished. Our work over the last three years has focused on the origin of such functional differences.

Lactotrope Subtypes

We have identified two subpopulations of lactotropes in pituitary cultures derived from adult male rats. We used the reverse hemolytic plaque assay, an immunological technique that permits the microscopic visualization of single-cell secretions. In this assay, cells releasing the appropriate hormone induce lysis of indicator erythrocytes. The size of the zone of hemolysis, or plaque, around an individual secretor provides an index of the cumulative amount of hormone released by that cell. In keeping with previous observations by Jimmy Neill (University of Alabama) on the secretory behavior of female rat lactotropes, we found that under basal conditions some prolactin-secreting cells form small plaques (SP lactotropes) and others induce large plaques (LP lactotropes). Thus SP and LP lactotropes secrete distinct amounts of prolactin per unit of time in the basal state.

Calcium Channel Activity as a Determinant of Lactotrope Heterogeneity

There is much evidence that calcium ion plays a major role in pituitary cells as an intracellular messenger in hormone secretion. In particular, basal prolactin secretion is thought to be sustained by calcium influx through plasma membrane calcium channels, which transiently open during spontaneous action potentials. This suggests that differences in basal secretory rate among lactotrope subtypes might arise, at least in part, from a differential expression of calcium channels. To test this possibility, we investigated the calcium channel activity of lactotrope subtypes, using electrophysiological techniques.

Our analysis indicates that both SP and LP lactotropes express two classes of voltage-gated calcium channels in the plasma membrane: low-voltage-activated (LVA) and high-voltage-activated (HVA) channels. The activity of LVA channels does not significantly differ between the two lactotrope subtypes. By contrast, the surface density of HVA channels is markedly higher in LP cells than in SP cells.

In addition, we tested the effect of nifedipine on prolactin secretion. Nifedipine is a dihydropyridine drug that selectively blocks HVA calcium channels in many types of excitable cells, including prolactin secretors. We found that nifedipine inhibits prolactin secretion by preferentially suppressing the LP lactotropes. In fact, a large proportion (around 50 percent) of LP lactotropes behave functionally as SP lactotropes in the presence of nifedipine. Our results, taken together, indicate that calcium entry through dihydropyridine-sensitive HVA calcium channels contributes to the high basal rate of prolactin secretion in LP lactotropes.

Sodium Channels in Lactotropes

We have characterized an additional difference in ion channel activity between the two lactotrope subtypes: membrane depolarization induces larger whole-cell sodium currents in LP lactotropes than in SP lactotropes. Such differences are not related to cell-to-cell variations in the kinetic properties of the sodium currents and persist after current amplitude is normalized by cell capacitance, which eliminates membrane

area as a variable. Thus sodium channels, like HVA calcium channels, are not uniformly expressed among SP and LP lactotropes.

Sodium channels should be functionally important for the secretory activity of lactotropes, as they favor the triggering of action potentials and thereby promote the opening of HVA calcium channels. Indeed, we have found that blocking the sodium channels of cultured cells with external tetrodotoxin drastically decreases the total amount of prolactin secreted. Furthermore, population analysis of prolactin plaque sizes suggests that tetrodotoxin, like nifedipine, preferentially inhibits prolactin secretion from LP cells.

These findings raise the question of what factors regulate ion channel expression in pituitary cells. They also renew our interest in the differentiation of lactotrope subtypes in the developing pituitary gland. Our current research is focused on these topics.

Ionic Channels in Sea Urchin Sperm Physiology

Alberto Darszon, Ph.D.—International Research Scholar

Dr. Darszon is Professor of Biochemistry at the Biotechnology Institute, National Autonomous University of Mexico, Cuernavaca, and Adjunct Professor of Biochemistry at the Center for Research and Advanced Studies, National Polytechnic Institute, Mexico City. He received his undergraduate degree in chemistry from the Universidad Iberoamericana in Mexico City, and his Ph.D. degree in biochemistry from the Center for Research and Advanced Studies. His postdoctoral research was done with Mauricio Montal at the University of California, San Diego. His honors include the National Science Award of the Mexican Academy of Scientific Research and both Guggenheim and World Health Organization fellowships.

THE molecular mechanisms involved in cell communication are at the forefront of research in biology today, since they play a key role in determining the behavior of organisms. For the past decade the main goal in our laboratory has been to understand the vital egg signals that allow sperm to fuse with the egg and fertilize it.

The sea urchin has proved to be an excellent model in which to study reproduction. From this animal enormous quantities of sperm (10^{10}/male) can be collected that respond to environmental stimuli rapidly, synchronously, and in a compulsory order. It has been shown that the flow of ions through the plasma membrane of sea urchin sperm participates crucially in the events leading to fertilization. Indeed, these are excitable cells that quickly respond to components from the outer layer of the egg, the jelly, with changes in their plasma membrane permeability.

Ionic fluxes play a fundamental role in the activation of respiration and motility, in chemotaxis, and in triggering the sperm acrosome reaction (AR). This latter reaction occurs within seconds after sea urchin sperm reach the egg and interact with its jelly coat. The AR involves important morphological changes that allow sperm to fuse with the egg. Participating in AR induction is an increase in the uptake of Ca^{2+} and Na^+ and efflux of K^+ and H^+. We would like to learn how these fluxes are related and elucidate the molecular mechanisms that orchestrate them to trigger the AR.

Previous results indicated indirectly that some ion fluxes occur through ionic channels. These are integral membrane proteins capable of forming hydrophilic pores through the membrane that allow the passive diffusion of ions at high rates (10^{6-7} ions/s). We have used model membranes (planar bilayers) formed from sperm components, and patch-clamp techniques in whole cells, to detect for the first time the activity of single channels in the plasma membrane of sea urchin sperm. These techniques, together with studies of membrane potential, intracellular Ca^{2+} ($[Ca^{2+}]_i$), and intracellular pH (pH_i) in whole sperm, have established the presence of K^+, Ca^{2+}, and Cl^- channels in this cell and are allowing us to explore their participation in the AR.

These findings have led us to propose a working hypothesis of how egg jelly-induced changes in ionic permeability might trigger the AR in sea urchin sperm. A receptor (to the inducing egg factor) in the plasma membrane of the sperm cell opens a Ca^{2+} channel, which deactivates in a few seconds. The activation of this channel modulates the opening of a second Ca^{2+} channel by poorly understood mechanisms that include a Ca^{2+}-dependent change in pH_i. At the same time or immediately after the first Ca^{2+} channel opens, a K^+ channel is activated, hyperpolarizing the cell and activating a voltage-dependent Na^+/H^+ exchange that increases pH_i. This latter change in pH_i is linked to the opening of the second Ca^{2+} channel and to a large depolarization.

In light of this working hypothesis, we thought it should be possible to hyperpolarize sperm artificially and induce an increase in $[Ca^{2+}]_i$ after a depolarization and AR. Valinomycin-induced hyperpolarization of sperm from the sea urchin *Lytechinus pictus* in K^+-free sea water raised pH_i, caused a small increase in $^{45}Ca^{2+}$ uptake, and triggered some AR. When the cells were depolarized with 30 mM KCl 40–60 seconds after the hyperpolarization, the pH_i decreased and there was a significant increase in $^{45}Ca^{2+}$ uptake, $[Ca^{2+}]_i$, and AR. Therefore the jelly-induced hyperpolarization may lead to the intracellular alkalinization required to trigger the AR and may modulate, on its own or via pH_i, Ca^{2+} channels involved in this process.

The sea urchin sperm offers distinct advantages over more complex cell types as a basic model for chemotaxis. As the spermatid matures, many intracellular organelles and macromolecules not involved in fertilization are eliminated. Sperm are incapable of division and devoid of the machinery for genome expression. However, they retain systems for sensing, swimming toward, and fusing with the egg, as described above.

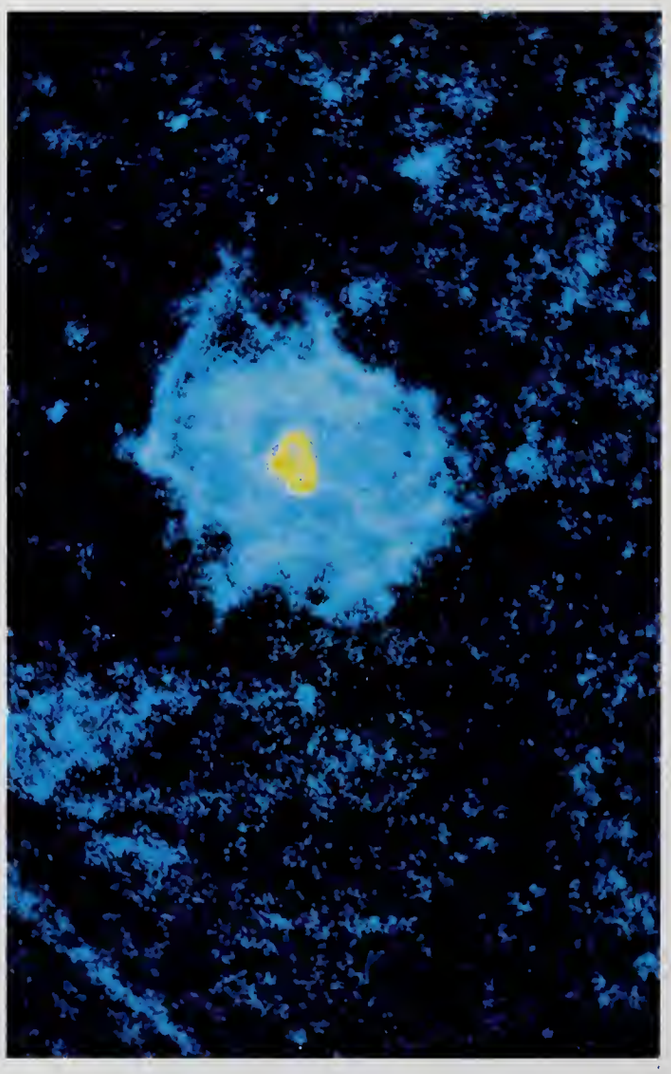
The sperm of the sea urchin *Arbacia punctulata* is attracted at nanomolar concentrations to a

small peptide isolated from the homologous egg. This peptide, called resact, is species specific and induces in sperm a transient increase in $[Ca^{2+}]_i$ and cGMP and an increase in pH_i. The egg of *Strongylocentrotus purpuratus* contains a peptide, named speract, that also elevates cGMP and $[Ca^{2+}]_i$ and induces an alkalinization and a K^+-dependent hyperpolarization.

Sea urchin sperm, however, are tiny cells (head diameter ~2 μm). This has precluded a careful characterization of their electrophysiological properties that would shed light on the molecular mechanisms determining their fascinating egg-induced behavioral changes. Recently, in collaboration with Donner Babcock and Martha Bosma from the University of Washington in Seattle, we found that it is possible to swell sea urchin sperm in diluted sea water. The swollen cells are spherical (~4 μm in diameter), immotile, and metabolically active, and they can regulate their $[Ca^{2+}]_i$, pH_i, and membrane potential.

The swollen cells respond to pM concentrations of speract with an increase in K^+-selective permeability that lasts for many seconds. We found that this permeability change, as well as the changes in $[Ca^{2+}]_i$ and pH_i that are seen at higher speract concentrations, also occurs in nonswollen sperm. An advantage of swollen sperm is that they can be much more easily patch clamped and single-channel-activity recorded. We have observed that pM speract activates a small K^+ channel. Thus swollen sperm provide new avenues to study ionic channels and their regulation by egg factors and second messengers.

This work is supported by grants from the Mexican Council for Science and Technology, the World Health Organization, and the Miguel Aleman Foundation.



Confocal immunofluorescent micrograph of Salmonella typhimurium *(yellow) interacting with a cultured epithelial cell and causing rearrangement of epithelial actin filaments (blue) around the invading bacterium.*

Research and photograph by B. Brett Finlay.

Host-Pathogen Interactions in Microbial Pathogenesis

B. Brett Finlay, Ph.D.—International Research Scholar

Dr. Finlay is Assistant Professor in the Biotechnology Laboratory and the Departments of Biochemistry and Microbiology at the University of British Columbia, Vancouver. He is a member of the Canadian Bacterial Diseases Center of Excellence. After receiving his B.Sc. and Ph.D. degrees in biochemistry from the University of Alberta, Edmonton, he conducted postdoctoral work on microbial pathogenesis in the laboratory of Stanley Falkow at Stanford University as a fellow of the Alberta Heritage Foundation for Medical Research.

IN all cases of bacterial disease, the bacterium or a bacterial product interacts with host cells or surfaces in either of two ways. It may adhere to the cell surface or may actually enter the cell and grow within (intracellular pathogen). Pathogens residing in an intracellular environment are protected from the host's immune systems, antibiotics, and other therapeutic agents. They often use this safe niche to multiply before disseminating to other sites and deeper tissue.

Our laboratory uses a multidisciplinary approach to define the interactions that occur between pathogenic bacteria and host cells. Essential to continuation of these interactions is the exchange of signals between pathogen and cell. We are studying the molecular nature of these signal transduction events in an effort to determine the mechanisms involved. Knowledge of these mechanisms should point the way to novel therapeutic strategies.

Several bacterial pathogens are used in these studies, since each organism has its individual features. Additionally, comparison of the signals generated by different pathogens can tell which mechanisms are common and which are unique.

Salmonella typhimurium: A Model for Intracellular Parasitism

For several reasons, *S. typhimurium* provides an excellent model for the study of intracellular parasitism. *Salmonella* species continue to cause significant health problems in both developed and less developed countries. These organisms have the capacity to enter into, survive, and replicate within host cells—features that contribute to virulence. Since *S. typhimurium* is a close relative of the nonpathogenic *Escherichia coli*, we have been able to use established molecular genetic techniques to study several aspects of *Salmonella* pathogenesis. Finally, *S. typhimurium* infection of the mouse closely mimics human typhoid fever.

Since *Salmonella* species (and many other pathogens) interact with intestinal epithelial cells following oral ingestion, we have been able to utilize monolayers of polarized epithelial cells grown on permeable substrates to study the interactions. These systems have several properties in common with columnar epithelial cells of the intestine. *Salmonella* species have the capacity to enter and penetrate through these polarized epithelial monolayers. Moreover, transposon mutants that are defective for such penetration have been isolated from *S. typhimurium*. Molecular characterization of the genetic loci that are disrupted by these transposons will provide clues about the bacterial products required for cell entry and penetration.

Once inside a vacuole within an epithelial cell, *S. typhimurium* finds itself within a presumably quite different environment. Very little is known about the microenvironment of any intracellular pathogen. To probe and define aspects of the vacuolar habitat, we have been using a bacterial reporter gene (*lacZ*) fused to several *S. typhimurium* genes that are variously regulated. For example, one can measure expression of reporter genes that are affected by oxygen levels, carbon source, pH, or iron or magnesium concentrations. These studies are providing clues about the nature of this intracellular niche. Additionally, we have begun to search for other *S. typhimurium* loci that are only induced when the bacterium is inside host cells, in the hope of finding other regulated genes.

After initial survival inside a host cell, *S. typhimurium* begins to multiply within its intracellular vacuole. Virtually nothing is known about its requirements there. We have identified three bacterial genes that are needed for *S. typhimurium* to multiply within host cells but not without. Characterization of these genes and other experiments may provide information about bacterial products necessary for intracellular growth.

Signal Transduction Between Host and Pathogen: Involvement of the Host Cytoskeleton and Tyrosine Kinases

Most intracellular pathogens require participation of the host cell for successful pathogen internalization. Functional actin filaments are often involved in bacterial uptake, and *S. typhimur-*

ium is one pathogen that requires host actin filaments for invasion. When *S. typhimurium* enters cultured epithelial cells, there is a large rearrangement in intracellular actin and the functionally related cytoskeletal proteins α-actinin, tropomyosin, and talin that surround the membrane engulfing the invading organism. This rearrangement, presumably needed for functional internalization, is triggered by bound extracellular bacteria and occurs in the region beneath adherent organisms. Once the pathogen is inside a vacuole, the epithelial cytoskeleton returns to its normal distribution.

We have been examining the signals transduced through the host cell membrane that are responsible for triggering this cytoskeletal rearrangement, and host cell signals that are necessary for the uptake of other pathogenic bacteria into epithelial cells. Host tyrosine kinases appear to participate in the uptake of several pathogens, including *Yersinia* species (causative agent of gastrointestinal problems), *Listeria monocytogenes* (a gram-positive organism associated with meningitis and serious infections of neonates), and enteropathogenic *Escherichia coli* (causative agent of diarrhea). We are using various kinase inhibitors, monoclonal antibodies that recognize phosphotyrosine residues, and radiolabeling of host phosphate pools to begin to identify the components of the signal transduction pathways that these bacteria pirate for their own use.

An *in Vitro* Blood–Brain Barrier to Study Bacterial Meningitis

Bacterial meningitis (infection of the brain lining) is a common and serious disease in both children and adults. In an effort to define more clearly the molecular mechanisms mediating this disease, we developed a model *in vitro* blood–brain barrier (BBB). This model uses primary isolates of microcapillary endothelial cells from bovine brain. Reseeded onto permeable substrates, these cells form impermeable monolayers with tight junctions and have several features characteristic of the BBB.

When *Haemophilus influenzae* (the most common cause of pediatric meningitis) is added to these monolayers, they are completely disrupted by a cytotoxic mechanism—similar to the effects observed *in vivo*. We have found that this cytotoxicity is mediated by bacterial lipopolysaccharide (LPS) in concert with a soluble serum factor. Current research has a twofold aim: first, to block this cytotoxicity with monoclonal antibodies against the serum factor in a primate meningitis model and second, to define the signal transduction pathways that are triggered in the endothelial cell by the LPS. In addition, the role of cytokines in this cytotoxic event is being studied.

Collectively this work provides several insights as to how pathogenic bacteria manifest disease. Studies with *S. typhimurium* have afforded new perceptions into the molecular biology of intracellular parasitism. Definition of the signal transduction pathways of invasive pathogenic bacteria is providing evidence that many pathogens exploit existing pathways, often utilizing them to pirate the host cytoskeleton for their own advantage. Work with the BBB has offered new opportunities for therapeutic intervention in treating meningitis.

Mechanisms of Transcriptional Regulation

Jack Greenblatt, Ph.D.—International Research Scholar

Dr. Greenblatt is Professor in the Banting and Best Department of Medical Research and the Department of Molecular and Medical Genetics at the University of Toronto. He earned his undergraduate degree in physics from McGill University, Montreal, and his Ph.D. degree in biophysics from Harvard University, where he studied bacterial gene regulation with Walter Gilbert. He pursued postdoctoral studies with Alfred Tissieres at the University of Geneva. He has received the Ayerst Award of the Canadian Biochemical Society.

THE ultimate focus of transcriptional regulatory mechanisms is RNA polymerase, a complex multisubunit enzyme whose activity is guided by interactions with a myriad of regulatory proteins and regulatory sequences in DNA or RNA. Much of our research is devoted to the basic enzymology of initiation and termination of transcription and to identifying and characterizing some of the key protein-protein interactions involved. In addition, we are examining how model regulatory proteins interact with the basic transcriptional apparatus. This work is also supported by grants from the Medical Research Council of Canada and the National Cancer Institute of Canada.

Initiation of Transcription by Human RNA Polymerase II

A set of general transcription factors (TFIIA, B, D, E, F, H) is necessary for RNA polymerase II to initiate the transcription of protein-coding genes. TFIID is the general factor that recognizes TATA sequences present in the promoters of many genes. After TFIID binds to the DNA, TFIIA and TFIIB recognize and bind to the TFIID-DNA complex. Subsequently the other general factors and RNA polymerase II assemble into a multiprotein complex at the promoter. We use protein affinity chromatography as a technique to identify direct protein-protein interactions involved in the assembly of this complex.

TFIID has been highly conserved during evolution. TFIID molecules from most or all eukaryotes, including fungi, insects, and plants, can function in transcription reactions containing RNA polymerase II and other general factors of human origin. By using yeast TFIID as a ligand for affinity chromatography, we have identified three human polypeptides that interact with TFIID and constitute human TFIIA. Curiously, TFIID columns do not retain TFIIB, nor do TFIIB columns retain TFIID, suggesting that a conformational change in TFIID induced by DNA binding facilitates its interaction with TFIIB.

By using RNA polymerase II as a ligand for affinity chromatography, we identified human RAP30 and RAP74, the small and large subunits of TFIIF. Human cDNAs encoding RAP30 and RAP74 have both been cloned, the latter in collaboration with Zachary Burton (Michigan State University). Using recombinant RAP30, we found that RAP30 is the subunit of TFIIF that binds RNA polymerase II. RAP30 prevents RNA polymerase II from associating with and transcribing nonpromoter sequences in DNA, a property also of bacterial σ factors. Indeed, we found that RAP30 can bind to *Escherichia coli* RNA polymerase and be displaced by the major bacterial σ factor, known as σ^{70}.

RAP30 also has a central role in promoter recognition by RNA polymerase II. In collaboration with Danny Reinberg (University of Medicine and Dentistry of New Jersey), we found that RAP30 can recruit RNA polymerase II to a preinitiation complex containing TFIIA, TFIIB, and TFIID. In fact, recognition of a promoter containing a TATA sequence by RNA polymerase II can be achieved with recombinant TFIIB, RAP30, and TBP, the TATA sequence–binding subunit of TFIID, all produced in *E. coli*. These three general factors, therefore, constitute a minimal set of proteins necessary and sufficient for promoter binding by RNA polymerase II. However, this preinitiation complex containing RNA polymerase II will not initiate transcription unless supplied with RAP74, TFIIE, TFIIH, and other factors. The roles, subunit compositions, and protein-protein interactions of some of these factors remain to be identified.

Regulation of Initiation by RNA Polymerase II

Many transcriptional activator proteins have two domains. One binds to regulatory sequences in DNA, and the other, known as an activation domain, provides an activating signal to the basic transcriptional apparatus. How these activation domains function is a fascinating question in regulatory biology. It has been the major focus of collaborative studies with my colleague, C. James Ingles (University of Toronto).

Many activation domains are highly acidic. A particularly potent one is found in the *Herpes*

simplex virus protein VP16. By using the VP16 activation domain as a ligand for affinity chromatography, we found that VP16 interacts directly with the TBP subunit of TFIID. Mutations in VP16 that reduce gene activation by VP16 also reduce its binding to TBP, and preliminary work has identified a mutation in TBP that reduces its ability to interact with and respond to VP16. Interaction of VP16 with TBP alters the association of TBP with the promoter. Therefore, VP16 may influence the ability of TBP to recruit other general factors and RNA polymerase II to the promoter.

A potent acidic activation domain is also found in p53, the product of a human anti-oncogene (a tumor-suppressing gene) that is mutated in about half of all human cancers. Like VP16, the p53 activation domain also binds TBP. By using affinity chromatography to search for other proteins that interact with p53, we recently discovered a new protein that interacts with the p53 and VP16 activation domains. At least some oncogenic mutations in p53 prevent binding of this protein to the p53 activation domain. The precise role of this novel protein in transcription is still under investigation.

Transcriptional Antitermination in *E. coli*

The N protein of bacteriophage λ prevents termination by RNA polymerase during transcription of λ operons expressed immediately after infection of *E. coli* cells. We have reconstituted antitermination by N *in vitro* in a system containing seven purified proteins: *E. coli* RNA polymerase, the *E. coli* transcription termination factor Rho, N, and four *E. coli* cofactors for antitermination (NusA, NusB, S10, and NusG). By using protein affinity chromatography and other methods, we have identified many of the protein-protein interactions in this system. Three factors, NusA, NusG, and S10, all bind to RNA polymerase. NusA is important for termination of transcription at some terminators. Since N binds to NusA, NusA is also an adapter that couples the antitermination factor N to RNA polymerase. In addition, NusB binds to S10 and NusG to Rho factor. In fact, NusG aids termination by Rho *in vitro* and is essential for termination by Rho *in vivo*.

The stable association of N with RNA polymerase requires an N utilization site (*nut* site) in the transcribed DNA and all four bacterial cofactors. The *nut* site RNA in the growing RNA transcript is recognized and bound by the proteins. N recognizes the *boxB* component of the *nut* site RNA, while NusB and S10 recognize a better version of the *boxA* component of *nut* site RNA, which is found in the antiterminator sequences of bacterial ribosomal RNA operons. Since mutations in RNA polymerase can prevent the formation of a stable multiprotein complex and can prevent binding of protein to the *nut* site RNA during transcription, we infer that a stable ribonucleoprotein complex containing *nut* site RNA and five antitermination proteins assembles on the surface of RNA polymerase. This extensively modified RNA polymerase can then transcribe through kilobases of λ DNA containing many transcriptional terminators.

The human immunodeficiency viruses (HIV-1 and HIV-2) produce antitermination factors, known as Tat, that recognize regulatory sequences in viral RNA, known as TAR. The human host factors involved in antitermination by Tat are unknown. Antitermination of HIV transcription by Tat may well resemble antitermination of λ transcription by N in many respects.

Ionic Homeostasis in White Blood Cells

Sergio Grinstein, Ph.D.—International Research Scholar

Dr. Grinstein is Head of the Division of Cell Biology at the Research Institute of the Hospital for Sick Children, Toronto, and Professor of Biochemistry at the University of Toronto. He received his Ph.D. degree at the National Polytechnic Institute, Mexico City, where he studied with David Erlij. His postdoctoral training was in two stages: initially at the Hospital for Sick Children under the supervision of Aser Rothstein, and later at the Federal Institute of Technology in Zurich with Giorgio Semenza. He has received the Ayerst Award of the Canadian Biochemical Society.

LEUKOCYTES constitute the body's first line of defense against invading microorganisms. These white blood cells first detect and engulf bacteria and other microbes, then secrete lytic enzymes and synthesize reduced oxygen metabolites to kill them.

These microbicidal processes call for pronounced changes in the generation and intracellular distribution of acid equivalents. First, the phagocytic vacuole wherein the microorganisms are trapped becomes markedly acidic. This evidently promotes the activity of the lytic enzymes released into the phagosome and may also facilitate its fusion with the vesicles containing bactericidal agents. Then too, the leukocytes' rate of metabolic acid production increases greatly during infection, threatening the stability of the cytosolic compartment, which must remain slightly alkaline to preserve optimal cell function.

The purpose of our research is to understand the mechanisms underlying phagosomal acidification, the pathways responsible for excess metabolic acid during leukocyte activation, and particularly the processes involved in the maintenance of the cytoplasmic pH under both resting and activated conditions.

Antiports and Channels in the Regulation of Cytosolic pH

The pronounced metabolic burst that leukocytes undergo when confronted by microorganisms or their products can be largely attributed to activation of an otherwise quiescent enzyme, the NADPH oxidase. The one-electron reduction of oxygen catalyzed by this enzyme is accompanied by oxidation of NADPH to $NADP^+$ and release of protons. Regeneration of NADPH through the hexose monophosphate shunt is a source of further proton production. If uncompensated, the proton production by these pathways would produce a massive intracellular acidification, incompatible with normal leukocyte function and possibly even their viability.

Three primary pathways appear to be involved in proton (equivalent) extrusion in activated leukocytes. The first and perhaps most important is an electroneutral exchanger (antiport) that transports protons out of the cells in exchange for extracellular sodium. A major isoform of this antiport has been identified in fibroblasts as a 110-kDa membrane glycoprotein. We have found that the antiport is active not only after stimulation but also in resting cells. Its activity, however, is greatly enhanced following the addition of bacterial chemoattractants or of molecules that mimic events in the intracellular signaling cascade triggered by microorganisms.

Our current and future research efforts in this area deal with the molecular characterization of the antiport in leukocytes and of the mechanisms underlying its activation during infection and inflammation. We are particularly interested in the subcellular localization of the antiports before and after stimulation, in the mechanisms whereby chemoattractants signal activation, and in the segregation and/or inactivation of antiports in compartments where sodium/proton exchange activity is counterindicated (such as the phagosome).

We have recently detected a second pathway that appears to be important in the extrusion of protons from activated leukocytes, namely a proton conductance, possibly a channel. This conductive path is essentially undetectable in quiescent cells but becomes clearly apparent when the cells are stimulated. Preliminary data indicate that this putative channel is present in neutrophils, macrophages, and the human leukemic cell line HL60. The conductive proton pathway could serve two important functions in the stimulated leukocyte: it could contribute to the extrusion of net acid equivalents from the cell, and it could also serve as a source for counterions to neutralize the voltage generated by the NADPH oxidase, proposed to be electrogenic.

Little is known at present about the conductive pathway. We are interested in defining its molecular identity, physiological significance, intracellular distribution, developmental pattern, and the molecular basis of its activation. In this regard, it is noteworthy that activation of the conductance closely mirrors the behavior of the

NADPH oxidase. We are in the process of exploring the relationship between these events, with particular interest in whether the putative channels are a component of the oxidase complex or whether assembly of the latter is required for activation of the conductance.

For this purpose, we have initiated studies using cells from patients with chronic granulomatous disease. These cells are defective in specific components of the NADPH oxidase. In future experiments we will attempt to detect the conductance electrophysiologically and to reconstitute its activity in cell-free systems. If a linkage between the two processes is established, purified or recombinant components of the oxidase will be used.

Proton Pumps in Cytoplasmic and Organellar pH Regulation

Recent experiments have also suggested that a third mechanism of proton extrusion is operational in stimulated leukocytes. Pharmacological evidence indicates that this latter pathway may be a proton-pumping ATPase of the vacuolar type. As described for the other systems, the activity of the pumps becomes clearly apparent after stimulation. At present, neither the subcellular location of the pumps is known nor the mechanism of activation understood. We will try to determine whether activation results from post-translational modification of pumps present in the relevant membrane or whether translocation between inactive and active compartments occurs. Immunochemical and molecular biological means will be used to localize the pumps and identify the type(s) of isozyme involved.

Proton pumps are also seemingly responsible for phagosomal acidification. We are currently interested in the source of the pumps that underlie this process and their mode of activation. We are also planning to study the determinants of the internal pH of the phagosome and other endomembrane compartments. Our current evidence suggests that differential counterion permeability, which was claimed to be the main source of pH heterogeneity, is not an important factor dictating intraorganellar pH. Differential pH sensitivity of the pumps, due to varying subunit composition, or a regulated proton leak permeability are our preferred hypotheses.

These studies are expected to contribute to our understanding of immune cell function and intracellular pH regulation in these and other cells.

Genetic Basis of Multidrug Resistance

Philippe Gros, Ph.D.—International Research Scholar

Dr. Gros is Associate Professor of Biochemistry at McGill University, Montreal, and a member of the McGill Cancer Center and the McGill Center for the Study of Host Resistance. He received his Ph.D. degree from McGill University and pursued postdoctoral training in molecular biology and cancer research at Massachusetts General Hospital with Joel Habener and at the Massachusetts Institute of Technology with David Housman.

TUMOR cells *in vivo* and *in vitro* can develop simultaneous resistance to a wide range of structurally and functionally unrelated cytotoxic drugs. Such multidrug resistance (MDR) severely impedes the chemotherapeutic treatment of many types of tumors. Structural or functional characteristics common to drugs of the MDR spectrum are few. In general these drugs are small, hydrophobic natural products that often contain a basic nitrogen atom and penetrate the cell by passive diffusion across the membrane. MDR is associated with a decreased intracellular drug accumulation and concomitant increased drug efflux from resistant cells, both ATP-dependent.

MDR is caused by the overexpression of a high-molecular-weight membrane phosphoglycoprotein called P-glycoprotein (P-gp). P-gp has been found capable of binding photoactivatable analogues of ATP and cytotoxic drugs, suggesting that it functions as an ATP-driven efflux pump that reduces the intracellular accumulation of drugs in resistant cancer cells. Recent studies have shown that increased P-gp expression in neuroblastomas and soft-tissue sarcomas causes lack of response to chemotherapy and is associated with very poor prognosis and outcome of these diseases.

P-gp is encoded by a small family of closely related genes, termed *mdr* or *pgp*, that share considerable sequence homology and common ancestral origins. This gene family has three members in rodents (*mdr1*, *mdr2*, and *mdr3*) and two in humans (*MDR1* and *MDR2*). We have isolated and characterized full-length cDNA clones corresponding to the three mouse genes and have deduced the amino acid sequences of the three predicted polypeptides. P-gps share considerable sequence homology (80–85 percent identity) and common structural features, including 12 predicted transmembrane domains and two nucleotide binding sites.

Each P-gp is formed by two homologous halves that show sequence conservation with a large group of bacterial transport proteins participating in the import and export of specific substrates in *Escherichia coli*. This evolutionary conservation is in keeping with P-gp's proposed drug efflux function.

The normal physiological function of P-gps has yet to be elucidated. Each P-gp isoform is expressed in a tissue-specific fashion, generally on the apical surface of secretory epithelial cells such as those of the bile canalicular, the brush border of the intestine, and the proximal tubule of the kidney. It has also been found in endothelial cells of the blood–brain barrier and in early pluripotent stem cells of the hematopoietic system. From these findings it appears that P-gp either plays a normal detoxifying role against environmental xenobiotics or transports normal physiological substrates yet to be identified.

Recently it was shown that the *mdr* gene family is itself part of a larger family of sequence-related genes shown to play key physiological functions in normal cells and tissues. These include the *STE6* gene of the yeast *Saccharomyces cerevisiae*, responsible for the transmembrane transport of the "a" mating pheromone; the *pfmdr1* gene of the malarial parasite *Plasmodium falciparum*, associated with chloroquine efflux from resistant isolates of this parasite; and in humans, the *CFTR* chloride channel gene, mutations of which cause cystic fibrosis, and the *RING* family genes, which code for peptide pumps implicated in antigen presentation by T lymphocytes. Therefore it appears that the *mdr* supergene family codes for membrane-associated transport proteins that may transport different types of substrates by the same mechanism.

We have carried out functional analyses of individual members of the mouse *mdr* gene family. For this, we have transfected and overexpressed cDNAs that correspond to each member of the family. We observed that *mdr1* and *mdr3*, but not *mdr2*, could directly confer MDR to otherwise drug-sensitive cells. In addition, the profile of drug resistance conferred by *mdr1* and *mdr3* appeared distinct.

One of the key unanswered questions about P-gp and MDR is how a single transport protein can apparently recognize and transport a large group

of structurally and functionally unrelated compounds. The identification of P-gp segments and residues implicated in drug recognition and transport is a necessary prerequisite to the rational design of new cytotoxic compounds capable of blocking or bypassing the action of P-gp in drug-resistant tumor cells. We have exploited the high degree of sequence similarity and striking functional differences detected among members of the mouse *mdr* family to identify, in chimeric and mutant proteins, segments and residues important for drug recognition. To identify the domains of *mdr1* that are essential for MDR and that may be functionally distinct in the biologically inactive *mdr2*, we have constructed 11 chimeric molecules in which discrete domains of *mdr2* have been introduced into the homologous region of *mdr1*. We have analyzed these chimeras for their capacity to confer MDR.

The two predicted ATP-binding sites of *mdr2* were found to be functional, as either could complement the biological activity of *mdr1*. However, the replacement of either the amino- or the carboxyl-terminus transmembrane (TM) domain region of *mdr1* by the homologous segment of *mdr2* resulted in inactive chimeras. Replacement of as few as two TM domains of *mdr1* from either the amino- or carboxyl-terminal halves by the corresponding segment of *mdr2* was sufficient to destroy *mdr1*'s activity. These observations suggest that the functional differences detected between *mdr1* and *mdr2* reside within the TM domains of the two proteins.

P-gps encoded by mouse *mdr1* and *mdr3* confer distinct drug resistance profiles. While both clones confer comparable levels of resistance to vinblastine (VBL), *mdr3* confers preferential resistance to actinomycin D (ACT), and *mdr1* to colchicine (COL).

To identify protein domains implicated in the preferential drug resistance encoded by either parental *mdr* clone, homologous protein domains were exchanged in a series of 16 hybrid cDNA clones, and the drug resistance profiles encoded by the corresponding chimeric proteins were analyzed. While all chimeric clones conferred similar levels of VBL resistance, the levels of ACT and COL resistance conferred by the various clones were heterogeneous, being either similar to the parental *mdr1* or *mdr3* clones or, in many cases, intermediate between the two.

Only those chimeric proteins carrying segments that overlapped both the amino and carboxyl sets of TM domains of the respective parent conveyed the parent's full preferential drug resistance profile. These results suggest that the resistance profiles encoded by *mdr1* or *mdr3*, possibly representing sites of drug-protein interactions, involve several determinants associated with TM domains from both homologous halves of P-gp. Recently we have tentatively identified one of these sites. We have observed that a simple serine-to-phenylalanine substitution at position 941 (*mdr1*) or 939 (*mdr3*), within predicted TM11, had a dramatic effect on the overall activity of the two pumps.

The modulating effect of this mutation on *mdr1* and *mdr3* varied for the drugs tested. It was very strong for COL and adriamycin (ADR) but only moderate for VBL. For *mdr1*, the serine-to-phenylalanine replacement produced a unique mutant protein that retained the capacity to confer VBL resistance but lost the ability to confer ADR or COL resistance. These results suggest that ADR-COL and VBL may have distinct binding sites on P-gp and that the serine residue within TM11 plays a key role in P-gp's recognition and transport of the former drugs. We are currently probing drug–P-gp interactions at this residue, using additional mutants together with COL and ADR analogues modified at key positions on their respective backbone.

Other investigators have found the same residue mutated in the *pfmdr1* gene from isolates of the human malarial parasite *Plasmodium falciparum* that are resistant to chloroquine (CLQ). Such resistance is caused by an increased ATP-dependent CLQ efflux and is associated with mutant alleles of the *mdr* homologue *pfmdr1*, mapping near TM1 or within TM11. Taken together, these and our studies indicate that the TM11 domain of *mdr* and *mdr*-like genes is critical for drug recognition and transport. Besides their high degree of hydrophobicity, no significant homology is detected between TM11 domains of *mdr1–3* and *pfmdr1* proteins. Both, however, have the potential of forming amphipathic helices.

The *mdr* $Ser^{939-941}$ residues and the mutated residues in *pfmdr1* fall within the hydrophilic side of this helix, and the mutant residues map near what appears to be the boundary of the hydrophilic side. These amphipathic helices may be important for the recognition of hydrophobic compounds, such as MDR drugs, that readily partition within the cell's lipid bilayer.

Control of Bacterial Protein Synthesis During Viral Infection

Gabriel Guarneros Peña, Ph.D.—International Research Scholar

Dr. Guarneros is Professor of Genetics and Molecular Biology at the Center for Research and Advanced Studies, National Polytechnic Institute, Mexico City. He received his undergraduate and M.Sc. degrees in microbiology, chemistry, and biochemistry in Mexico City, and his Ph.D. degree in molecular biology from the University of California, Berkeley, where he studied with Harrison Echols. He joined the staff of the Center after doing postdoctoral work in molecular biology at the University of Geneva, Switzerland, in Harvey Eisen's laboratory. He has been awarded fellowships from the Guggenheim Memorial Foundation, the Commission of the European Communities, and the Sistema Nacional de Investigadores, Mexico.

INFECTING viruses divert the functions of the host cell for their own development. To achieve this, the viral genome directs the synthesis of regulatory molecules, which reorient cell functions to conform to the specific viral development program. There is abundant evidence of transcriptional control during phage λ infection of *Escherichia coli*, where the phage genome directs the synthesis of regulatory proteins that alter the cells' transcriptional pattern. Little is known, however, about the translational control in λ-infected cells. Our laboratory has pursued a case of translational control involving λ RNA sequences, named *bar*, and peptidyl-tRNA hydrolase (Pth), a bacterial enzyme essential for protein synthesis.

The Target in the Cell

Bacterial mutants partially defective in the activity of Pth are unable to support the growth of λ phage. We have shown that the corresponding mutations are located in the *pth* gene region. This result was further confirmed by sequence analysis; the mutations are base substitutions within a translational open reading frame that corresponds to the *pth* gene. Pth, the gene product, was isolated and characterized chemically and enzymatically.

Pth hydrolyzes peptidyl-tRNAs to yield free tRNAs and peptides. It has been proposed that the enzyme is a scavenger of peptidyl-tRNAs that have dropped off the ribosomes during editing of misincorporated amino acids in polypeptide chains. The Pth function is essential for the cell, as inferred from the fact that a heat-sensitive mutant of Pth accumulates peptidyl-tRNAs and stops protein synthesis upon shift to the nonpermissive temperature. Moreover, the enzyme is ubiquitous among organisms from bacteria to mammals.

The possible role of Pth in ribosome-bound hydrolysis of peptidyl-tRNA at the step of polypeptide chain termination has not been supported by the work of others. Preliminary results, however, indicate that Pth may be involved in polypeptide chain termination (see below), and we aim to investigate this possible participation.

Nature of the λ Regulator

We have isolated phage mutants that overcome the bacterial Pth defect. These mutations defined several genetic sites named *bar*. DNA sequence analysis of two of these, *bar*I and *bar*II, revealed that the mutations affect nearly identical 16-bp segments having dyad symmetry.

The inhibition of phage development by mutants defective in Pth requires transcription through wild-type λ *bar* regions. Transcripts themselves, not polypeptides, seem to be the active molecules. A plasmid system in which short λ *bar* sequences were cloned in front of an active promoter somehow mimics the λ inhibition effect.

Transcription of wild-type *bar* in plasmids is lethal to Pth-defective (but not wild-type) cells. This effect is specific, because constructs carrying mutant *bar* sequences are not lethal. Transcription through a vector harboring a synthetic nucleotide sequence as short as 20 bp mimicking *bar*I caused *pth* mutant lethality; therefore we think that the 16-bp *bar* sequence is the core of the inhibitory transcripts. Protein synthesis is shut off soon after *bar* transcript induction, but RNA synthesis continues for several hours. Thus the lethal effect is probably caused by a rapid inhibition of protein synthesis.

How does *bar* RNA inhibit protein synthesis? Preliminary data on nonsense codon-specific suppression, obtained in collaboration with Emanuel Murgola (M. D. Anderson Cancer Center, Houston), have led us to propose a unifying model that implicates *bar* RNA and Pth in peptide chain termination. The model postulates, first, that *bar* RNA can interfere with the termination of UGA-mediated translation by antiparallel base-pairing with ribosomal 16S RNA, and second, that mutant Pth causes a defect in polypeptide termination facilitating *bar* RNA–16S RNA interaction.

We are attempting to test this model through a direct assay of *bar* RNA interaction with cell components and the effect of Pth on polypeptide chain termination. We will also investigate possible genes and gene products that interact with Pth through the isolation of second-site mutations that suppress *pth* defect.

Biochemistry of Pth and Regulation of *pth* Expression

To purify Pth protein, we have exploited the fact that cells harboring *pth* plasmids overproduce the wild-type enzyme. Part of our current efforts are devoted to overproducing and purifying mutant enzymes to compare their biochemical properties. Mutant Pths may have different patterns of specificity for various amino acyl-tRNAs. Also our laboratory is working on the control of *pth* expression by analyzing the transcriptional and translational properties of the gene.

What Does *bar* Do for λ?

We have considered two possible roles for *bar* regulation in λ biology, taking into account the two functions proposed for Pth. First, *bar* RNA may control the relative levels of specific tRNAs to fit the profile of codon usage in λ. This function, perfectly tolerated in normal cells, may lead to inhibition of phage development or to lethality in cells defective for Pth. Second, *bar* RNA may act on termination and/or initiation of polypeptide chains directed by phage transcripts. Among λ genes, UGA is the most frequent termination codon (not UAA, as in the host), and in the λ genome it is not uncommon for a gene's termination codon to overlap partially the initiation codon of the next gene, producing the sequence AUGA. The fact that the *bar* core RNA contains the sequence AUGA, and the alleged interaction of this sequence with ribosomal 16S RNA, suggest a role for these overlapping genes in polypeptide chain-termination (-initiation) events.

Molecular Genetics of Photosynthesis and Carbon Assimilation in Plants

Luis R. Herrera-Estrella, Ph.D.—International Research Scholar

Dr. Herrera-Estrella is Professor and Head of the Department of Plant Genetic Engineering at the Center for Research and Advanced Studies, National Polytechnic Institute, Irapuato. He received his undergraduate degree as biochemical engineer from the National Polytechnic Institute in Mexico and his Ph.D. degree from the State University of Ghent, Belgium, where his thesis advisors were Marc Van Montagu and Jeff Schell. The subject of his thesis was the expression of foreign genes in plant cells. As a postdoctoral researcher in Ghent, he studied the regulation of light-inducible plant genes and the movement of proteins into the chloroplast. His honors include the Minoru and Ethel Tsutsui Distinguished Graduate Research Award in Science from the New York Academy of Sciences and the Javed Husain Award for Young Scientists from UNESCO.

PHOTOSYNTHESIS and carbon assimilation are the most important biochemical and molecular events in the life cycle of higher plants and, indeed, are key to the provision of nutrients for the entire food chain. Solar energy is first collected in the chloroplasts of photosynthetic tissues, mainly of leaves, by light-harvesting antennas composed of chlorophyll and protein molecules. The collected energy is then used to convert atmospheric carbon dioxide (CO_2) into triose phosphate molecules. These three-carbon molecules proceed through a series of reactions, called the Calvin-Benson cycle, that culminates in the production of sugars from which all the organic molecules required for plant life are synthesized.

More specifically, triose phosphate molecules are converted in the cytoplasm of photosynthetic, or source, cells into sucrose, which is translocated through the phloem to feed the nonphotosynthetic, or consumer, tissues (i.e., roots, flowers, seeds, tubers). Assimilated carbon is stored temporarily or permanently in the form of starch in both source and consumer tissues. The starch in seeds or tubers provides most of the carbon and energy for the germination and development of new plants.

The light-dependent production of ATP and NADPH, the reductive assimilation of CO_2, and sucrose and starch synthesis are interlinked and interdependent. These processes must be coordinated *in vivo* at both the biochemical and genetic level (i.e., at the level of gene expression). The balance between the efficiency of CO_2 fixation, sucrose translocation and uptake, and assimilation of sucrose in consumer tissues plays a fundamental role in determining the productivity of any given plant species. This balance is affected by both genetic determinants of the individual and its interaction with the environment.

Our laboratory is interested in studying the molecular events that control the biochemical processes involved in carbon assimilation in plant cells. One aspect that we are investigating is how light regulates genes involved in photosynthesis. Ribulose 1,5-bisphosphate carboxylase oxygenase (RuBisCO) is a multimeric enzyme composed of eight identical small subunits (SS) and eight identical large subunits (LS). In the so-called C_3 plants, RuBisCO carries out the initial CO_2 fixation step, utilizing the five-carbon sugar ribulose bisphosphate to produce two three-carbon derivatives (triose phosphate molecules). The gene family encoding the SS is located in the nuclear genome; the gene encoding the LS is located in the plastid genome. How genes located in different cellular compartments are coordinately regulated (those for both SS and LS are regulated by light and should produce the corresponding subunits in equimolar amounts), and how the active RuBisCO enzyme is assembled from its subunits, are some of the questions we wish to answer.

In collaboration with June Simpson, we showed previously that the 5′-flanking sequences of genes encoding the small subunit of the RuBisCO (*ss* genes) and the chlorophyll *a/b*–binding proteins (*cab80* gene) are responsible for the light-inducible transcription of these genes. We have also shown that enhancer and silencer elements are involved in the tissue-specific and light-inducible expression of the *cab80* gene. The cis- and trans-acting elements involved in the regulation of the *cab80* gene are being analyzed. From these studies a 7–base pair repeated element within the 5′-flanking region of the *cab80* gene has been identified as the site of interaction with a putative regulatory DNA-binding protein.

Many attempts to assemble RuBisCO *in vitro* from its isolated subunits have failed, suggesting that the plant cell *in vivo* provides other requisite components. It has been suggested that RuBisCO assembly requires the participation of proteins that act as molecular chaperones to promote the correct interaction between the subunits, ensuring the assembly of the catalytically active enzyme. Two 60-kDa proteins termed

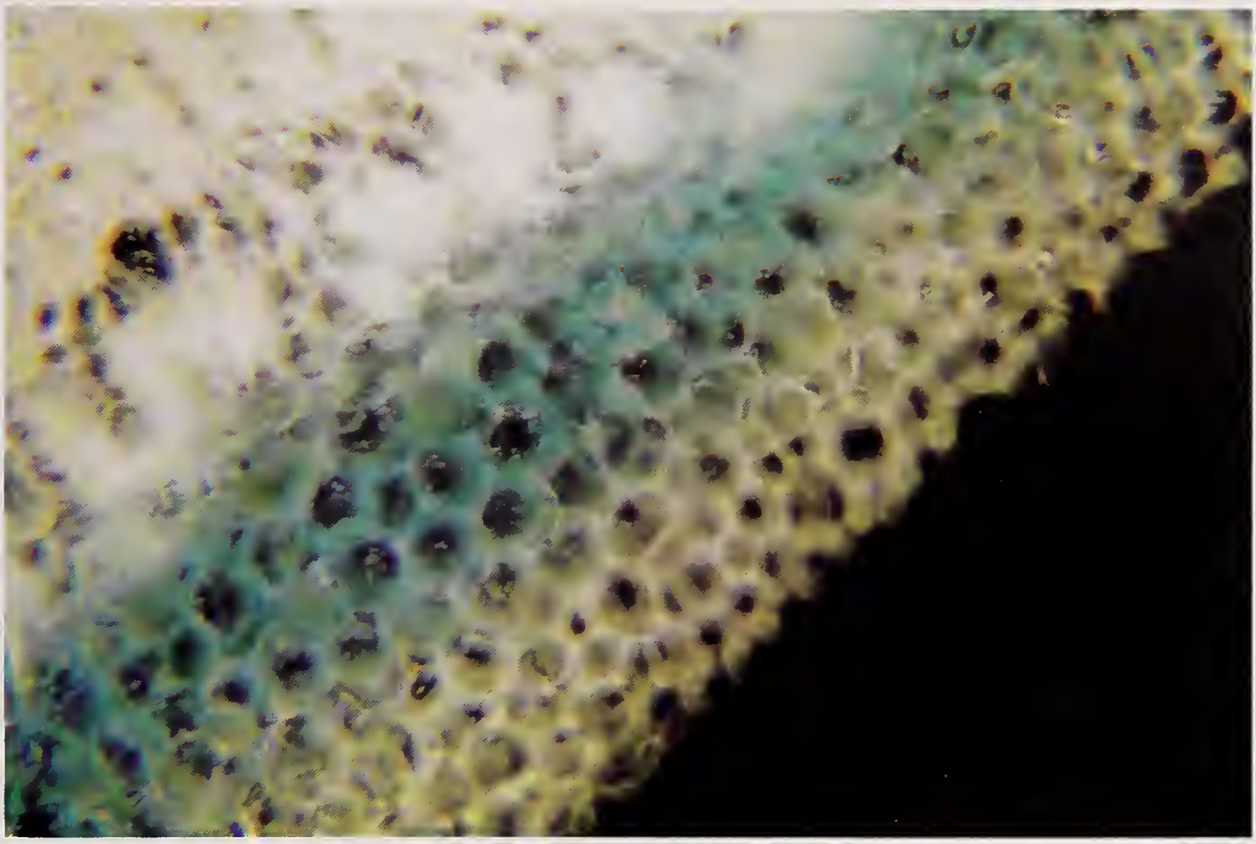
chaperonins (Cpn60) α and β have been suggested as participating in RuBisCO assembly. Since the availability of functional RuBisCO is important for carbon assimilation, we have isolated two cDNAs and a genomic clone for Cpn60 β.

To study whether Cpn60 β is indeed involved in RuBisCO assembly, and whether it is specific for RuBisCO or plays a more general role in protein assembly, we are currently generating transgenic plants containing a chimeric gene to produce an antisense RNA for Cpn60 β. The RNA should interact with the Cpn60 β mRNA and arrest the production of Cpn60 β polypeptides, thus producing transgenic plants that will allow us to assess the role of Cpn60 β in RuBisCO assembly. We are also constructing chimeric genes composed of the promoter of a Cpn60 β gene and the coding sequence of the β-glucuronidase bacterial reporter gene, to study the tissue-specific and environmental regulation of this gene in transgenic plants.

Again, the main assimilates from photosynthesis are triose phosphate molecules that are converted into sucrose to be translocated to consumer tissues. Several physiological and biochemical studies indicate that sucrose 6-phosphate synthase (SPS) is the limiting enzyme for this conversion. To study the role of SPS in carbon assimilation, and the interrelation of the genes encoding this enzyme with photosynthetic genes, we are currently isolating SPS cDNA clones, using monospecific SPS antibodies and polymerase chain reaction technology in collaboration with Horacio Pontis from Mar del Plata–Argentina. (The work on SPS is supported in part by a grant from the Rockefeller Foundation.)

Opposite: Light microscopy of transverse stem sections of transgenic tobacco plants harboring a chimeric gene in which the β-glucuronidase–coding sequence is under control of the chaperonin 60 β promoter. The blue staining indicates the cell-type–specific expression directed by the promoter at the basal (A) and apical (B) stem regions of transgenic plants.

Research and photograph by Eduardo Zabaleta in the laboratory of Luis Herrera-Estrella.



A

B.

Expression of the mouse En-2 *gene in a band of cells across the midbrain-hindbrain junction. The 10-day transgenic embryo contains a* lacZ *reporter construct expressed from* En-2 *DNA regulatory sequences. The cells expressing* lacZ *(dark) are visualized by histochemical staining of the* lacZ *gene product.*

From Sedivy, J., and Joyner, A. 1992. Gene Targeting. *New York: W.H. Freeman, p 161.*

Gene Pattern Expression in Early Embryogenesis

Alexandra L. Joyner, Ph.D.—International Research Scholar

Dr. Joyner is a Senior Scientist at the Samuel Lunenfeld Research Institute of Mount Sinai Hospital, Toronto, and Associate Professor of Molecular and Medical Genetics at the University of Toronto. She received her B.Sc. degree in zoology and her Ph.D. degree in medical biophysics from the University of Toronto. She did postdoctoral work in mammalian development with Gail Martin at the University of California, San Francisco.

THE establishment of the basic body plan requires an intricate coordination of cell-cell interactions that appear to be controlled largely by the genetic program handed down from generation to generation in our DNA. Many of the genes that run the program of pattern formation have been identified in *Drosophila* and shown by mutant analysis to regulate the development of embryonic regions rather than the differentiation of cell types. In keeping with this, many of these genes are expressed early in embryogenesis in spatially defined patterns.

The primary focus of research in my laboratory has been to identify and study mouse homologues of *Drosophila* pattern-formation genes. This work has been based on the premise that a conservation of gene structure through evolution reflects a corresponding conservation of gene function. Work over the last few years in many laboratories, including mine, has shown this to be the case.

A second research project has involved developing and applying a new type of random screen, called the gene trap, for genes expressed in a spatially defined manner during early mouse embryogenesis.

Mouse Homologues of the *Drosophila* Gene *engrailed*

The fruit fly body is divided into a number of repeated units referred to as segments, and the *engrailed* (*en*) genes are known to be required for proper development of the posterior half of each segment. The *en* gene has characteristics of a "switch" that can direct cells down a posterior, as opposed to anterior, developmental pathway. Molecular characterization of the gene's product has shown it to be a transcription factor that binds DNA through a motif called a homeodomain. We have been studying *En-1* and *En-2*, the mouse homologues of the *Drosophila en* gene, to determine whether they also act as transcription factors controlling pattern formation.

One striking difference between these homologous mouse and fruit fly genes is that the fly *en* gene is expressed during embryogenesis in 14 stripes (one for each segment), whereas the vertebrate *En* genes are first expressed in a single band across the developing mid- and hindbrain junction. *En-1* and *En-2* continue to have a spatially defined expression pattern in the brain until the cells begin to differentiate into particular cell types. *En* expression then switches and becomes cell-type specific.

Many other mouse homologues of *Drosophila* pattern-formation genes have also been analyzed. A recurring theme is expression in spatially defined patterns early in development, particularly in the nervous system. This suggests that at least part of the mechanism for laying down the basic body plan, and especially for specifying different regions of the nervous system along the anterior-posterior axis, involves the coordinate expression of different sets of these developmental switch genes.

One of our objectives has been to make transgenic mice that lack the *En* genes. To date we have made mice that are deleted for the *En-2* homeodomain-coding DNA sequences. These mice do not show a major disruption of the whole mid- and hindbrain region but show a distinct disruption in the pattern of folds in the cerebellum. We are now studying the developmental progression of the defect to determine the cellular basis of the mutant phenotype. That the cerebellum, uniquely among adult brain structures, expresses *En-2* in the absence of *En-1* indicates a functional redundancy between *En-1* and *En-2*. Thus both *En* genes must be deleted before their function in development of the mid- and hindbrain can be studied.

A second aspect of our *En* research is to identify other genes in the same genetic pathways as *En-1* and *En-2*. One of our approaches has been to seek transcription factors that regulate *En* expression. As a first step, we have analyzed the DNA sequences around the *En-1* and *En-2* genes and have identified fragments that will direct expression of the *lacZ* reporter gene to the mid- and hindbrain junction. These fragments can now be further subdivided and used to identify the transcription factors that bind these sequences and regulate *En* expression.

In addition, we are using these DNA transcription regulators to express other pattern-formation genes aberrantly in the mid- and hindbrain region. In this way we can test, for example, whether other homeodomain-containing genes, expressed in more-posterior regions than *En*, can switch the developmental program of mid- and hindbrain cells to more-posterior regions.

We are also interested in identifying genes that are regulated by *En*. One approach we are taking is to analyze mouse homologues of *Drosophila* genes that are regulated by *en*. To this end, we have cloned homologues of the *Drosophila cubitus interruptus* (*Ci*) gene. The mammalian genes are called *Gli*, since one member of the family was found to be overexpressed in gliomas. We can now analyze the expression of the *Gli* and other potential *En*-regulated genes in mice lacking *En-2*.

Screening for Genes Expressed in Spatially Defined Patterns

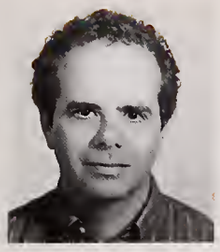
With a view to identifying other types of genes involved in pattern formation, we devised a screen that takes advantage of mammalian gene structure and mouse embryonic stem (ES) cells. The screen involves randomly integrating a vector we refer to as a gene trap into the ES genome. The vector has a splice acceptor site upstream of the reporter gene *lacZ*. We reasoned that if the vector integrated in the correct orientation into the intron, then splicing of the gene would be directed to the reporter gene. We have now confirmed this by cloning and characterizing a number of transcripts for *lacZ*-endogenous gene fusions.

Cells expressing *lacZ* can be revealed with a histochemical stain. ES cells expressing the reporter are then reintroduced into a host embryo for *lacZ* expression pattern analysis. We have now analyzed the expression pattern of the cloned endogenous genes from two *lacZ* insertions and have shown that the *lacZ* pattern faithfully mimics that of the endogenous gene.

Finally, *lacZ* insertions that show a spatially defined expression pattern can be transmitted into transgenic mice and the insertions analyzed for mutant phenotypes. Two out of three such mice showed mutant defects. Thus we have demonstrated the feasibility of the gene trap approach and are now carrying out a larger screen for candidate pattern-formation genes.

All of this work is also funded by grants from the Medical Research Council and the National Cancer Institute of Canada, the National Institutes of Health, and Bristol-Myers Squibb.

Diagnostic Use of RNA Replication in Infectious Diseases



Paul M. Lizardi, Ph.D.—International Research Scholar

Dr. Lizardi is Professor of Biochemistry at the Biotechnology Institute, National Autonomous University of Mexico, Cuernavaca. He received his Ph.D. degree from the Rockefeller University and conducted postdoctoral research in embryology with Donald Brown at the Carnegie Institution of Washington, Baltimore. After serving as Associate Professor at Rockefeller as an Andrew Mellon Foundation fellow, Dr. Lizardi spent a sabbatical year at Massachusetts General Hospital, Boston, and held a visiting professorship in genetics at Harvard Medical School.

INFECTIOUS diseases are frequently managed by health professionals without definitive identification of the pathogen. Classical approaches to diagnosis, such as direct microscopic observation, cultivation, or infection of susceptible hosts, are often too slow or cumbersome for routine medical care. Modern laboratory tests based on the use of antibodies provide effective tools for diagnosis of a number of infectious diseases, but most antibody tests are designed to detect the presence of a host immune response rather than the pathogen itself. Thus there is a need for techniques permitting the rapid and reliable detection of infectious agents so that epidemiological monitoring and patient management can be more effective.

A direct way to detect a pathogen is to identify its genetic material, which invariably contains unique sequence patterns. However, the genetic material is usually so minute in a biological sample that its detection presents an extraordinary technical challenge. Molecular biologists have been up to the task, and techniques developed recently permit the generation of millions of copies of DNA or RNA segments in the test tube by a process of exponential amplification. The best-known amplification method is the polymerase chain reaction (PCR), in which DNA strands are sequentially separated by heating and then copied with DNA polymerase. The PCR limit of detection is about 50 molecules of target, which is over 100,000 times as sensitive as a typical enzyme-linked immunoassay.

Amplified RNA Binary Probes

My laboratory is developing alternative amplification methods for the detection of RNA or DNA in biological samples. The work is being carried out in close collaboration with Fred Kramer at the Public Health Research Institute in New York City and Jack Szostak at Massachusetts General Hospital in Boston. Our methods exploit several interesting properties of RNA molecules:

- Relatively short molecules of RNA (15–30 nucleotides) have long been known to form stable helical structures when bound to perfectly complementary single strands of DNA or RNA. The thermodynamic stability of these helical complexes distinguishes them from similar complexes containing mismatched bases. Hence one can design an RNA probe that will bind very strongly, and uniquely, to a segment of the DNA or RNA of an infectious agent in a biological sample.
- RNA probe molecules can be joined to other RNA molecules in a target-dependent manner. That is, joining will only take place if the molecules are aligned on a complementary target strand that serves as a guide for the joining event. We catalyze the joining by an enzyme called a ribozyme ligase, isolated in Jack Szostak's laboratory. The ligase is itself an RNA, derived from a naturally occurring intron called group I, originally discovered in *Tetrahymena thermophyla* by Thomas Cech (HHMI, University of Colorado at Boulder).
- A specific class of molecules known as replicatable RNAs can be produced exponentially under natural conditions, so that millions of copies are generated in minutes. An enzyme called RNA replicase catalyzes reactions in which RNA single strands, parent and daughter, are forced apart during synthesis, in contrast to DNA-dependent reactions, in which the two strands remain annealed. The best characterized of these enzymes is Q-beta replicase. We have shown that replicatable RNAs, in the presence of this enzyme, can harbor RNA probe inserts without loss of replicative efficiency.

We have devised schemes in which probe binding, RNA joining, and exponential replication of the joined RNA are used in an assay to detect the presence of a specific target sequence from an infectious agent. Probes that are not joined are not replicated. The signal in these assays is replicated RNA, generated by joined probes and readily quantitated by fluorescence staining. The intensity of the signal is proportional to the number of targets present in the original sample.

However, a number of technical problems remain to be solved before optimal sensitivity and specificity can be achieved in these assays. For example, the ligation step involving the ribo-

zyme ligase is still relatively inefficient. (About 25 percent of the ligation-competent binary probe molecules are joined in a 1-hour incubation.) An actual assay would require a ligation efficiency closer to 85 percent. Fortunately, recent developments in RNA biochemistry suggest a possible solution to this problem.

Directed Evolution of Ribozymes

We propose to develop more-efficient ribozyme ligases for our assays by directed evolution *in vitro*. Rachel Green, in Jack Szostak's laboratory, has selected novel ribozymes with altered catalytic efficiency by Darwinian selection in a test tube. A Darwinian selection experiment begins with the synthesis of a large number of DNA molecules, each containing the sequence of a ribozyme. Known as a mutant pool, the population of molecules contains point mutations generated during chemical synthesis. Each molecule contains just a few mutations, located at random positions in the sequence, and thus the population contains over a trillion (10^{12}) variants. The enzyme T7 RNA polymerase is then used to generate RNA copies of the DNA, creating over a trillion different mutant ribozymes. These are briefly incubated under RNA ligation conditions, where they are given the opportunity to catalyze a chemical reaction that joins them to a special piece of RNA.

Those ribozymes that succeed in carrying out the ligation reaction are chosen as survivors in a subsequent step, while those that fail to participate in catalysis are lost by dilution. Survivors are allowed to increase in number by reverse transcription (which converts RNA into DNA), followed by PCR amplification of the DNA. This series of steps effectively rewards all competent, efficient ribozyme mutants by reconverting them into DNA and copying them many times.

At this point the DNA is transcribed again into RNA, and a second cycle of Darwinian selection begins, exactly as above. After four cycles of selection, a relatively small number of mutant ribozyme sequences will be present in the DNA population, instead of the original millions. These molecules are the fittest: they were able to catalyze ligation reactions in all four rounds of selection. The molecules are then sequenced in order to compare them to the original parental ribozyme.

Using the methods outlined above, Rachel Green isolated a number of interesting mutant ribozymes that display high efficiency in a specific type of ligation reaction. The experiments used to generate these ribozymes mimic natural variation and selection. While it takes nature hundreds, thousands, or even millions of years to evolve better enzyme catalysts, RNA enzymes can be improved in the laboratory by a Darwinian selection process in a matter of months. The term *in vitro genetics* has been coined to describe these man-made selection schemes. We are working on improvements that should enable us to generate ribozyme ligases better suited to our assays, in order to improve the process of joining RNA binary probes.

Multiplexed Diagnostic Assays

Epidemiological monitoring of infectious diseases is a complex undertaking in a developing country like Mexico. Diseases must be monitored both in large, modern urban settings, like Mexico City, and in sparsely populated, rural areas like the Chiapas countryside. The list of diseases causing significant mortality and morbidity includes AIDS, malaria, amebiasis, tuberculosis, hepatitis, typhoid fever, diverse intestinal infections, and more recently cholera. There is a need for developing accurate, low-cost technologies to detect pathogens in the field, and also a need for extremely efficient, high-tech tools for the epidemiological laboratory and the blood supply centers. A valuable epidemiological tool would be a diagnostic assay capable of detecting any of a number of pathogens in a single clinical sample. Such a tool is known as a multiplex assay.

Under the auspices of the Rockefeller Foundation and the World Health Organization, our institute has established a collaborative research program with Stanford University School of Medicine to develop and implement state-of-the-art techniques for epidemiological assessment in Mexico. The Stanford group is headed by Gary Schoolnik from the Division of Geographic Medicine and HHMI. Using a multiplex PCR assay, they have conducted a series of experiments demonstrating the simultaneous detection of any of several bacterial pathogens in diarrheal stool.

Assays based on RNA amplification could be multiplexed by using a combination of binary probes. Each binary probe pair would contain RNA sequences designed for binding to an individual pathogen, and as many as 10 probe pairs could be mixed together in a single assay. Eventually multiplexed assays could be automated, and cost-effectiveness would be very favorable because several candidate infectious agents would be assayed at once.

Lineage-Specific Gene Expression in Caenorhabditis elegans

James D. McGhee, Ph.D.—International Research Scholar

Dr. McGhee is Professor in the Department of Medical Biochemistry at the University of Calgary, Alberta. He received his B.Sc. degree in physiology and biochemistry from the University of Toronto and his Ph.D. in molecular biology from the University of Oregon. His postdoctoral research was done in the laboratory of Gary Felsenfeld at the National Institutes of Health. He is a Medical Scientist of the Alberta Heritage Foundation for Medical Research.

ONE of the most important problems in developmental biology is to understand the mechanism of lineage-specific gene expression. That is, how does a developing embryo manage to express a particular gene in one cell or cell lineage and not in the many other cells of its body? Furthermore, how does it tell time, in order to express this gene only at the correct point in development?

We are approaching these problems in the simple nematode, or roundworm, *Caenorhabditis elegans*. The reasons for choosing such a simple animal are that it comprises only about a thousand cells and, more importantly, that the division pattern and cell lineage of every one of these is known. We are studying the expression of genes in the *C. elegans* intestine, since this is the simplest lineage available.

When the developing embryo has only eight cells, one of these (called the E cell) is the progenitor of the animal's entire intestine. As a marker for biochemical differentiation of the gut, we have characterized a simple hydrolytic enzyme, a nonspecific carboxylesterase. As shown in the figure, the gene coding for this esterase (called *ges-1*, standing for gut esterase) is only expressed in the developing intestine. Presumably the natural function of the enzyme has something to do with the worm's digestion, but mutants that do not express esterase appear to be perfectly viable.

The *ges-1* gene is transcribed from the genome of the embryo at a point in development when the embryo has only 100–200 cells and the developing gut lineage consists of only four cells. Micromanipulation experiments have shown that *ges-1* is expressed normally in embryos in which non-gut cells have been removed or destroyed. In other words, we have no evidence for any interaction between gut and non-gut cells. An intriguing feature of *ges-1* expression is that it is completely dependent on DNA synthesis during the cell cycle in which the embryo has a total of eight cells, i.e., the cell cycle just after the gut has been clonally established.

The problem of understanding intestinal-specific expression of the *ges-1* gene is twofold. One must define regions in the DNA to which transcription factors bind, and then identify and characterize the transcription factors that bind to these regions. To address the first part of the problem, we have injected the cloned DNA from the *ges-1* gene back into a *ges-1*–nonexpressing mutant and have been able to reconstitute accurate gut-specific expression. We have used this transformation assay to identify sequences that appear to be necessary for correct gut expression.

We have also found regions, however, whose deletion causes the *ges-1* gene to be expressed not in the gut but in specific sets of cells in the pharynx, in the body wall musculature, and in the hypodermis. These cells belong to either the sister or cousin lineage of the gut.

The simplest model to explain these ectopic staining patterns invokes lineage-specific repressors that would normally keep the *ges-1* gene silent. When the repressor-binding site is deleted, the normally silent gene can be expressed. We are testing this model by attempting to identify short discrete sequences that cause lineage-specific repression of a gene that is otherwise ubiquitously expressed. We are also attempting to provoke ectopic expression of the *ges-1* gene by injecting large amounts of putative repressor-binding sites as a competitor.

There is a curious and unexpected feature of these ectopic *ges-1* staining patterns. With constructs that produce staining in pharynx, muscles, and hypodermis, the three staining patterns are by and large exclusive: individual embryos show only one of the patterns, and combinations of patterns are rare. Whatever the molecular explanation turns out to be, the phenomenon suggests that the embryo is making a decision very early in development and stably propagating this decision thereafter. Indeed, this decision must actually be made at or before the two-cell stage of the embryo, since this is the last point when the different lineages are common.

We are also working on the second part of the problem of lineage-specific gene expression, namely to identify the protein factors that bind to

the *ges-1* gene and control its expression only in the gut and only at the correct embryonic stage. Here one of the few limitations of *C. elegans* as an experimental organism becomes apparent: it is essentially impossible to obtain large amounts of synchronized embryos on which to do biochemical experiments. We are taking two routes past this obstacle.

The first route is to turn part of our efforts to *Ascaris suum*, the well-known parasite of pig intestines. *Ascaris* adults are millions of times larger than adult *C. elegans*, yet their embryos are remarkably similar. Furthermore, *Ascaris* females produce hundreds of thousands of fertilized eggs per day, which develop more-or-less synchronously. We hope to use this system to watch the arrival and departure of transcription factors during early nematode development.

The second route is based on our finding that *C. elegans* embryos can be blocked in mid-development by exposure to the chemotherapeutic agent fluorodeoxyuridine. This has allowed us to produce nuclear extracts from these blocked embryos and to detect a number of protein factors (possibly as many as a dozen) that bind to the 5′-flanking sequence of the *ges-1* gene. We have concentrated on two such factors.

The first factor binds to a DNA sequence that we have tentatively identified as a gut "activator." The protein is not detectable in oocyte cytoplasm and thus must be produced (transcribed or translated) sometime during early development. The second factor binds to a sequence that is completely conserved between *C. elegans* and the related nematode *Caenorhabditis briggsae*. Unlike the first factor, this second factor is present in the oocyte cytoplasm and must migrate at some point in early development into the embryonic nuclei. The second factor would thus be a candidate for a classic cytoplasmic "determinant" produced by the mother worm and then segregated into the gut lineage during early cell cycles.

If we can indeed describe the constellation of transcription factors responsible for activating the *ges-1* gene in the developing gut, will we really be much further ahead in understanding its control? In some respects, all this work will only move the problem one step further back in development, to the question of how the transcription factors themselves got where they did. However, one of the main advantages of *ges-1* as an experimental system is that we only have 3–5 cell cycles through which to regress, until we reach the oocyte. Thus it should be possible to describe the complete molecular logic that controls this simple gene.

This work was supported by the Medical Research Council of Canada and by the Alberta Heritage Foundation for Medical Research.

A newly hatched larva of the nematode Caenorhabditis elegans. *The dark red precipitate reflects the esterase activity of the* ges-1 *gene, expressed in all 20 cells of the gut.*

From Edgar, L.G., and McGhee, J.D. 1986. Dev Biol *114:109–118.*

Cytokine Regulation of Effector Functions in Immune Responses

Tim R. Mosmann, Ph.D.—International Research Scholar

Dr. Mosmann is Professor and Chair of the Department of Immunology at the University of Alberta, Edmonton. He received his Ph.D. degree in microbiology at the University of British Columbia, Vancouver, and obtained postdoctoral training in Toronto and Glasgow as a fellow of the Medical Research Council. After four years as Assistant Professor in Immunology at Alberta, he spent eight years as a senior scientist at DNAX Research Institute in Palo Alto. He then returned to the University of Alberta.

AFTER infection the host's immune system must respond by specifically recognizing the invading organism and inducing the effector mechanisms that will most effectively destroy the pathogen. Determination of the correct antigen-specific recognition structures is mediated by the process of clonal selection, whereby B or T cells are selectively activated if they already express receptors specific for the infecting agent. Choice of the appropriate effector mechanism is determined to a large degree by T cells and the cytokines they secrete.

The immune system can attack infectious agents by a number of means. Antibody is highly effective at neutralizing toxins and free virus and at coating bacteria to enhance their recognition and destruction by phagocytic cells. There are also various cell-mediated cytotoxic mechanisms, such as killer T cells, macrophages, and granulocytes, that are most useful against intracellular infectious agents. Cytokines secreted by activated T cells regulate the accumulation and activation of these cells at sites of infection, and some cytokines also have direct cytotoxic functions.

Recruitment and activation of macrophages and granulocytes occur strongly during a delayed-type hypersensitivity (DTH) reaction, which presents an effective response against intracellular infections. In contrast, antibody responses are normally more effective against extracellular pathogens. It has long been known that antibody and DTH responses are often "either-or" responses of the immune system, although the mechanism of this reciprocal regulation has only been partially resolved.

Several years ago we and others described two types of T cell that secrete very different patterns of cytokines. TH1 cells produce interleukin-2 (IL-2), interferon-γ (IFN-γ), and lymphotoxin (LT), whereas only TH2 cells produce IL-4, IL-5, IL-6, IL-10, and P600. These cytokines have multiple and profound effects on various aspects of the immune response, and so TH1 and TH2 cells have markedly different functions.

TH2 cells induce antibody production by B cells. In particular, a strong TH2 response is associated with high production of IgE, the antibody that causes allergy. Although TH1 cells can also induce antibody production under some circumstances, these cells are much more effective at inducing DTH. Thus the choice of TH1 or TH2 responses influences the balance between DTH and antibody responses.

Although the TH1 and TH2 patterns account for a major part of the immune response during certain infections, these probably represent extremes, and there are other T cells with different cytokine secretion phenotypes. These other patterns may be involved in the many possible types of immune response that occur against different infectious agents.

Since the type of immune response induced against a particular pathogen is often characteristic for that agent, the choice of effector function must be precisely regulated. It appears that cross-regulation by TH1 and TH2 cytokines plays a major role in determining the effector functions. Some of this cross-regulation is now understood: IFN-γ produced by TH1 cells can inhibit proliferation of TH2 cells, and IL-10 produced by TH2 cells inhibits the activation of TH1 cells.

Effector T cells such as TH1 and TH2 are probably derived from precursor cells that secrete only IL-2. After initial stimulation with antigen, these cells presumably differentiate into TH1, TH2, and other cells. This process is influenced by cytokines—e.g., IL-4 encourages the production of more cells that secrete IL-4. However, the full regulation of this process is largely unknown, and this is currently an active area of investigation in our laboratory and others. Since the differentiation process is quite rapid (e.g., a few days), we are developing single-cell cytokine detection methods so that we can follow differentiation of single clones of cells.

Two methods are currently possible: the mRNA for different cytokines can be measured in single cells after amplification by polymerase chain reaction, and cytokine protein secreted by single cells can be analyzed by a colorimetric spot assay using anticytokine monoclonal antibodies. Both

methods are feasible for detecting single cytokines produced by one cell, and we are currently adapting them to simultaneous measurement of multiple cytokines. Since there is considerable evidence of several different cytokine secretion phenotypes, 5–10 cytokines must be analyzed from each cell to obtain a true picture of T cell diversity. Once developed, this methodology will be applied to the differentiation of T cells during normal immune responses, allowing the signals that influence differentiation to be determined.

The characteristic cytokines of TH1 and TH2 are major determinants of the function of these cells. For example, IL-4, IL-5, and IL-10, produced by TH2 cells, contribute to the production of IgE, mast cells, and eosinophils—all associated with the development of allergy. IFN-γ is a major contributor to DTH reactions, inhibits many of the functions of IL-4, and inhibits TH2 cells directly. We have been interested for some time in additional cytokines of the two T cell types.

Recently we have been working on P600, originally discovered as a clone in a cDNA library of activated TH2 cells. P600 is particularly interesting, since it is expressed only after activation and is not produced by TH1 cells. The sequence of the open reading frame in the P600 cDNA clone encodes a small protein with a leader sequence characteristic of secreted or membrane proteins. Since all of these properties suggest that P600 may be an additional TH2-specific cytokine, we decided to test for possible functions of the P600 protein.

We transfected the P600 cDNA clone into monkey cells and observed secretion of a new protein of the expected size. After screening in a wide variety of biological assays, we found that P600 induces the production of large numbers of adherent cells from bone marrow. The functions of these cells are currently under investigation to determine how the effects of P600 fit into the overall pattern of TH2 functions during immune responses.

The different T cell cytokine secretion phenotypes were originally discovered in the mouse, and initially there was some doubt that similar phenotypes exist in humans. However, recent data on human T cell clones and immune responses have made it clear that the TH1 and TH2 patterns are also important in human diseases. Apparently several parasitic diseases strongly induce TH1- or TH2-biased responses, and only one of these responses is usually able to clear the infectious agent. Recent data indicate that these two cytokine patterns are also important in leprosy. Thus a knowledge of the regulation of these patterns will have considerable potential for the design of better immunomodulatory treatments and vaccines.

Cellular and Molecular Basis of Variability in Entamoeba histolytica

Esther Orozco, Ph.D.—International Research Scholar

Dr. Orozco is Professor of Genetics and Molecular Biology in the Department of Experimental Pathology at the Center for Research and Advanced Studies, National Polytechnic Institute, Mexico City. She received her bachelor's degree in chemistry and biology from the University of Chihuahua, Mexico, and her Ph.D. degree from the National Polytechnic Institute. She has been a visiting professor in several institutes around the world, including Harvard School of Public Health in Boston and the Weizmann Institute of Science in Israel. Dr. Orozco counts among her honors a Guggenheim Foundation Award and the Dr. J. Rosenkranz 1991 Award given by Syntex.

MORE than 500 million people throughout the world are infected by *Entamoeba histolytica*, the protozoan parasite responsible for human amebiasis. Ninety percent of those infected do not present clinical symptoms, while the rest develop colitis, dysentery, or hepatic abscesses.

In 1925 Brumpt proposed the possible existence of two *E. histolytica* species, one pathogenic (P) and the other nonpathogenic (NP), and he and others have continued to advance this possibility. There are data both supporting and opposing this concept. For instance, migration patterns of several enzymes in a high number of amebic isolates has led to the establishment of more than 20 *E. histolytica* zymodemes. Interestingly, the zymodemes of P and NP trophozoites differ.

According to the opinion of several groups, zymodemes and virulence are stable phenotypes. However, experimental evidence against the stability of zymodemes and virulence has also been documented. Resolution of this controversy is of fundamental importance clinically. If *E. histolytica* comprises two species, cyst passers spreading the NP form are not a public health menace, and asymptomatic carriers need not be treated. However, if the harmless *E. histolytica* can turn into a virulent one, a different decision may be indicated. In any case, the molecular basis of conversion, if it occurs, would be a highly interesting phenomenon to study.

We have focused this controversy by 1) investigating cloned *E. histolytica* populations after their isolation from asymptomatic carriers throughout the axenization process, 2) comparing *E. histolytica* clones with diverse origins, and 3) detecting and cloning DNA sequences present only in *E. histolytica* clones that express pathogenicity.

Study of NP *E. histolytica* Throughout Axenization

Cloned trophozoites (MAV-1), with an initial NP zymodeme I and isolated from an asymptomatic carrier, were cultured under polyxenic, monoxenic, and axenic conditions. Zymodeme I of polyxenic MAV-1 trophozoites, cultured in Robinson's medium, presented an NP zymodeme XII when trophozoites were cultured in Jones' medium. Monoxenic trophozoites grown in the presence of *Fusobacterium symbiosum* retained NP zymodeme I (with slow-running hexokinases [HKs]), an α-phosphoglucomutase (PGM) band, and lack of pathogenicity. Surprisingly, when MAV-1 trophozoites were grown under axenic conditions, the original NP zymodeme switched to a P zymodeme, with fast-running HKs and a β PGM. Virulence, defined as the trophozoites' ability to infect experimental animals and damage target cells, was also expressed.

The analysis of *E. histolytica* zymodemes indicates that P as well as NP trophozoites have a set of genes for a given enzyme. Excluding the fast-running HKs and the β PGM that have not been reported for NP isolates, all other isoenzymes are expressed in P and NP trophozoites, including slow-running HKs. A certain type of bacteria in the medium could be participating in the regulation of isoenzyme expression. We will continue our studies on the factors involved in the expression of isoenzymes in P and NP trophozoites, in order to look for conclusive evidence on zymodeme and virulence switching—evidence that should dissipate skepticism about conversion.

Differences in Genetically Related Clones

Functional and biochemical dissimilarities mark clones of diverse origin: P trophozoites show a high rate of phagocytosis, destroy cell culture monolayers, and produce hepatic abscesses in experimental animals. On the contrary, NP trophozoites generally do not show these virulence-related properties. Fortuitously, they express low virulence.

Monoclonal and polyclonal antibodies against certain antigens can distinguish between P and NP trophozoites. The presence of the major amebic cysteine protease involved in the cytopathic effect of trophozoites is also specific to P tropho-

zoites. However, the so-called P trophozoites form a heterogeneous group of *E. histolytica* isolates, which show virulence in a broad spectrum. This indicates that virulence-involved genes are differentially regulated.

On the other hand, amebic strains are constituted by trophozoites displaying different phenotypes. Clones from a given strain express virulence in the same broad spectrum as amebic strains. Some clones are avirulent (are they NP?); others are highly virulent, and specific monoclonal and polyclonal antibodies are able to discriminate between them.

The simplest interpretation for this is that strains are a mixture of genetically unrelated trophozoites living together in the human intestine. However, clones derived from cloned populations differ phenotypically and genotypically. The differences include virulence, antigens, and genetic divergence. Clones isolated from clone A (strain HM1:IMSS) show the P zymodeme but differ in virulence and in the expression of the 112-kDa adhesin, involved in both phagocytosis and virulence of *E. histolytica*.

Detection and Cloning of DNA Sequences Present Only in Pathogenic Trophozoites

Several *E. histolytica* genes show high similarity among P strains but differ from their homologous genes in NP strains. Molecular studies have given the strongest evidence to support the existence of two *E. histolytica* species. To date, there is no explanation for the genomic divergence shown by P and NP trophozoites. However, reticence to accept that there are two species based on their polymorphism is generated by the remarkable genomic plasticity of this parasite, even among clones derived from cloned populations.

Genomic divergence is showed by strains and clones cultured under different conditions. Genetically related clones obtained from a cloned population show polymorphism in several genes of the same clone cultured in different laboratories (clone A, strain HM1:IMSS). They also differ in their molecular karyotype and in the chromosomal location of some genes, indicating genome rearrangements. Finally, we have identified, cloned, and sequenced a 1.6-kilobase DNA fragment from clone A (virulent) that is altered in nonvirulent (NP?) clones derived also from clone A.

We are currently studying genetic mechanisms underlying this intriguing phenomenon. One hypothesis, highly speculative, is that there are genes in *E. histolytica* that, under pressure, are selectively rearranged, modified, or amplified before being transcribed. If certain genes were amplified in P trophozoites and not in NP, single-copy or slightly modified genes could not be detected in comparative experiments. If this were true, it would explain, at least in part, the polymorphism found in P and NP strains.

The presence of "cassettes" carrying different genes that could be amplified and expressed selectively is another possibility. We are trying to develop a genetic transformation system in *E. histolytica* to introduce foreign genes as landmarks to follow up the genome modifications that will explain the erratic behavior of this parasite, or at least the differences between P and NP trophozoites.

This work is supported by a grant from the National Council for Science and Technology, Mexico.

Phosphorylation and Protein-Protein Interactions in Signal Transduction

Tony Pawson, Ph.D.—International Research Scholar

Dr. Pawson is Senior Scientist at the Samuel Lunenfeld Research Institute of Mount Sinai Hospital, Toronto, where he holds the Apotex Chair of Molecular Oncology; Professor of Molecular and Medical Genetics at the University of Toronto; and a Terry Fox Cancer Research Scientist of the National Cancer Institute of Canada. He received his Ph.D. degree from the University of London while working at the Imperial Cancer Research Fund. Prior to moving to Toronto, Dr. Pawson did postdoctoral work at the University of California, Berkeley, and was a faculty member of the University of British Columbia.

MANY of the polypeptide hormones that control development in the embryo and cell growth, differentiation, and metabolism in the adult bind to the extracellular region of receptors that span the plasma membrane. These receptors, in their cytoplasmic region, contain a protein kinase domain that phosphorylates tyrosine residues. Normally these catalytic receptors are only active when engaged by the appropriate growth factor. However, structural alterations, resulting from cancer-causing mutations in the relevant genes, can render them active even in the absence of a growth factor. Such aberrantly active receptors can trigger unrestricted cancerous cell growth.

Activated growth factor receptors can elicit many changes in stimulated cells, including changes in gene expression, cellular architecture, cell-cell communications, and cellular metabolism. An important question involves the identities of the intracellular targets of receptor tyrosine kinases—targets that control signal transduction pathways within the cell.

SH2-containing Proteins Bind Activated Growth Factor Receptors

The first thing growth factor receptors do upon activation, apparently, is phosphorylate themselves on tyrosine. Receptor autophosphorylation serves as a switch to elicit the tight binding of intracellular signaling proteins to the receptor. These receptor-binding proteins include Ras GTPase-activating protein (GAP), phospholipase C-γ, phosphatidylinositol (PI) 3′-kinase, and Src-family cytoplasmic tyrosine kinases. They have the attributes expected of signaling proteins, in the sense that they regulate specific intracellular signal transduction pathways and in some cases are known to be phosphorylated, and activated, by receptor tyrosine kinases. How are these functionally diverse signaling proteins all able to recognize autophosphorylated receptors? They each contain one or two copies of a protein domain, the Src homology 2 (SH2) domain, which recognizes specific autophosphorylated sites. Thus tyrosine phosphorylation of the receptor itself creates a tight binding site for SH2 domains and promotes a strong interaction between the activated receptor and its target SH2-containing proteins.

There is evidently some specificity to these SH2-receptor interactions. Different SH2 domains bind tenaciously to distinct phosphorylated sites on a receptor, suggesting that the amino acid sequence surrounding the phosphotyrosine residue is important in determining target specificity.

A Network of SH2-mediated Interaction

An increasing number of SH2-containing proteins that can bind to activated growth factor receptors have been identified. These are all likely to regulate some aspect of intracellular signaling, though in several cases the relevant biochemical pathways have not been defined.

The association of SH2 domains with autophosphorylated growth factor receptors is one indication of their general ability to bind tyrosine-phosphorylated sites. As an example, considerable data suggest that the SH2 domain of the c-Src cytoplasmic tyrosine kinase interacts with an inhibitory tyrosine phosphorylated site in its own carboxyl-terminal tail, thereby repressing c-Src tyrosine kinase activity. However, once c-Src is enzymatically (and oncogenically) activated by removal of the carboxyl-terminal phosphorylation site, the SH2 domain has a positive transforming effect, probably by virtue of its ability to bind substrates of the kinase domain and participate in the formation of signaling complexes. Thus the c-Src SH2 domain has multiple functions. GAP is another example of a signaling protein whose SH2 domains can interact with several different phosphotyrosine-containing proteins.

We are pursuing the structure and function of SH2 domains, using a variety of biochemical, cellular, and genetic approaches.

Receptor Tyrosine Kinases in Embryonic Development

We have recently cloned cDNAs for two novel receptor tyrosine kinases, *elk* and *nuk*, which are primarily expressed in the mammalian nervous system. During formation of the nervous system in the early mouse embryo, *nuk* assumes a very specific and interesting pattern of expression. We are using genetic manipulation in the mouse to determine whether *nuk* is involved in controlling neural differentiation.

Dual-Specificity Protein Kinases

Protein kinases are typically divided into those that can phosphorylate serine/threonine residues and those that phosphorylate tyrosine. We have recently identified two mammalian protein kinases that can phosphorylate both serine/threonine and tyrosine when artificially expressed in bacteria. We are investigating the significance of this observation for mammalian cells.

One of these kinases, Clk, is related to members of the *cdc2* protein kinase family. The second, Nek, is similar to the product of the *nimA* gene, required for initiation of mitosis in the fungus *Aspergillus nidulans*. The identification of *nek* as a mouse relative of *nimA* suggests that this gene may have been conserved in evolution, as have other members of the cell cycle machinery. These observations prompt the speculation that control of the cell cycle involves protein kinases with dual specificity.

Additional funding for this work was provided by the National Cancer Institute of Canada and the Medical Research Council of Canada.

Chemical and Functional Characterization of Scorpion Toxins

Lourival Domingos Possani, Ph.D.—International Research Scholar

Dr. Possani is Professor and Chairman of the Department of Biochemistry at the Biotechnology Institute, National Autonomous University of Mexico, Cuernavaca. He received his B.S. degree in natural history from the Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, and his Ph.D. degree in molecular biophysics from the University of Paris, France. He completed the academic requirements for his Ph.D. degree in biochemistry at the Rockefeller University, working with Edward Reich on the isolation of the acetylcholine receptor.

SCORPIONISM is a serious public health problem in certain areas of the world. In Mexico alone, more than 200,000 people are stung by scorpions each year. Thanks to serotherapy with horse antisera, the number of deaths is kept to approximately 700–800 annually. Not all scorpions are dangerous to humans. Among the 136 species known to live in Mexico, only 6 present a real risk. They belong to the family Buthidae, genus *Centruroides*. The species dangerous to humans are *Centruroides elegans*, *C. infamatus infamatus*, *C. limpidus limpidus*, *C. limpidus tecomanus*, *C. noxius*, *C. sculpturatus* (which occurs also in Arizona), and *C. suffusus suffusus*.

Half of the casualties are due to one species, represented by two subspecies: *C. limpidus limpidus* from the states of Guerrero and Morelos, and *C. limpidus tecomanus* from the state of Colima. All dangerous species abide in the Pacific coast region. No dangerous scorpions are found in the Gulf of Mexico region, the Yucatan Peninsula, or Mexico City.

Envenomation by the sting of these arachnids is due to low-molecular-weight peptides called toxins. More than 100 such toxins have been purified and characterized. There are three well-known classes of scorpion toxins: short-chain peptides (38–39 amino acid residues long), which are K^+ channel blockers; medium-chain peptides (61–66 amino acid residues), specific for Na^+ channels; and long-chain peptides (about 70 amino acid residues), toxic to insects and crustaceans. Recently a fourth class seems to have emerged with the discovery of low-molecular-weight proteins (about 120 amino acid residues) that affect the Ca^{2+} release channel from sarcoplasmic reticulum.

The study of scorpion toxins is scientifically, medically, and biotechnologically interesting. They represent several families of structurally related polypeptides with exquisite preference for ion channel molecules, which makes them a model for future new drugs to control cell excitability or for new types of specific insecticides. And they represent excellent tools for studying cellular responses, permitting discrimination among receptor molecules in the membranes and investigation of various types of ion channels.

The contribution of our research group, which comprises a multidisciplinary interaction with investigators from several countries, began with the isolation and chemical characterization of toxin gamma, a potent Na^+ channel blocker, from the Brazilian scorpion *Tityus serrulatus*. Our work was extended at the beginning of the 1980s by the discovery of the first peptide capable of blocking the delayed rectifier K^+ channel from the squid axon. It is a short-chain peptide (39 amino acid residues) and was subsequently shown to affect other types of K^+ channels. This peptide was named noxiustoxin, after the scorpion *C. noxius*.

Our studies continued with the purification and characterization of more than 25 different peptides from scorpion venoms, specific blockers of Na^+ channels from a variety of different tissues. At least eight additional small-chain peptides similar to noxiustoxin were also isolated and characterized. To study the function of these toxins, we used neurotransmitter-release and -binding studies with brain synaptosomes and mainly electrophysiological studies with a variety of excitable cells. Our more recent progress in this field is focused on the three lines of research mentioned below.

Chemical Synthesis of Peptides and Monoclonal Antibodies

We have recently synthesized, by the solid-phase method of Merrifield, more than 100 different peptides corresponding to segments of the amino acid sequence of the scorpion toxins. These peptides and the preparation of a dozen distinct monoclonal antibodies allowed us to probe for specific structural regions of the toxins. For example, the nonapeptide at the amino-terminal region of noxiustoxin was shown to recognize and affect K^+ channels in a manner similar to the native peptide. These results are promising in the context of possible development of new drugs aimed at controlling cellular excitability through K^+ channels.

Using monoclonal antibodies, we found four main distinct antigenic determinants in the Na^+ channel target toxins of eight different species of scorpion from the American continent. One of the antibodies was shown to produce an important protection to experimental animals envenomated with highly purified toxin. This result is also important for developing possible synthetic peptides for vaccination.

Cloning of Scorpion Toxin Genes

The most recent advance of our laboratory, in collaboration with Francisco Bolivar and Baltazar Becerril, was the successful isolation of several clones that code for peptides corresponding to the amino acid sequence of scorpion toxins. A cDNA library was prepared from the telsons of the scorpions *C. noxius* and *C. limpidus limpidus*. In the future we hope to use directed mutagenesis to obtain specific modified toxins in order to study the structure-function relationship of these ligands with specific ion channels.

Search for New Toxins

New, unknown peptides toxic to insects and crustaceans were purified and are being sequenced. These are from several species of scorpion, including some not dangerous to humans. Also, in collaboration with Hector Valdivia and Roberto Coronado from Madison University, we are conducting a search for specific toxins that target the Ca^{2+} release channel from the sarcoplasmic reticulum.

This research is supported in part by grants from the Consejo Nacional de Ciencia y Technologia, Mexico, and Direccion General del Personal Academico, from the National Autonomous University of Mexico.

Venom extraction by electrical stimulation of a scorpion, species Hadrurus concolorous. *Photograph by Manuel Varela and Lourival Possani.*

Protein Crystallography in the Study of Infectious Diseases

Randy J. Read, Ph.D.—International Research Scholar

Dr. Read is Assistant Professor of Medical Microbiology and Infectious Diseases and of Biochemistry at the University of Alberta, Edmonton. He obtained his undergraduate and doctoral training in the Department of Biochemistry at the University of Alberta, then pursued postdoctoral training in protein crystallography with Wim Hol at the University of Groningen in the Netherlands, before returning to Edmonton.

INFECTIOUS organisms—viruses, bacteria, and parasites—must overcome many obstacles to cause disease. They must gain entry to their favored niche in the host, obtain nutrients, evade attacks of the immune system, and spread to new hosts. These are complex problems with complex and varied solutions. But the pathogenic mechanisms used by microorganisms to infect and cause disease are gradually coming to light, in some cases at the level of the key molecules. The interest in pathogenesis is more than academic, since the understanding of these processes can be exploited to prevent or treat disease.

Our interest is in advancing the understanding of infectious disease at the molecular level, using the technique of x-ray crystallography to study the three-dimensional structure of important proteins. There are two major aspects to our work. We are studying the crystal structures of a number of proteins that are involved in pathogenesis, some of which are described below, and we are developing methods to exploit this kind of structural information in the design of new drugs.

Bacterial Toxins

Many pathogenic bacteria produce toxins that cause cell and tissue damage and can be responsible for the most severe effects of the illness. Bacterial toxins often belong to the A-B class, having a two-part structure in which the B (binding) subunit binds to the surface of a target cell and the A (active) subunit enters the cell, carrying out the toxic action. We are involved in studying the crystal structures of pertussis toxin (in collaboration with Glen Armstrong at the University of Alberta, and with Connaught Laboratories in Toronto) and verotoxin (in collaboration with James Brunton, University of Toronto).

Pertussis toxin (PT) is produced by *Bordetella pertussis*, the bacterium that causes whooping cough. There is a major interest in the role of this toxin in improved vaccines. Currently, killed whole-cell vaccines are used for whooping cough. They are effective, but have an undesirable level of toxicity. The side effects should be reduced or eliminated in a defined vaccine produced from genetically engineered proteins. It has been shown that PT is a necessary component of effective whooping cough vaccines, but it must be rendered nontoxic for safe use. A three-dimensional structure of PT would help show how to remove its toxicity while preserving the surface features of the molecule recognized by the immune system. In addition, we would gain a better understanding of how the B component of the toxin recognizes the surface of target cells, how the A subunit enters the cell, and how it carries out its toxic action. Exploiting the intense radiation produced by synchrotrons, we have collected x-ray data from crystals of this toxin and from several potential heavy-atom derivatives, but the determination of the phases needed for the visualization of the structure is still in progress.

The Shiga toxin family is a group of closely related toxins produced by *Shigella dysenteriae* type 1 (Shiga toxin) and by certain strains of *Escherichia coli* (verotoxins, or VTs). Shiga toxin is associated with bacterial dysentery, a serious problem in many developing countries. The strains of *E. coli* that produce VTs cause a disease often referred to as "hamburger disease," because it can be acquired from contaminated hamburger. VTs can provoke the hemolytic uremic syndrome and are thereby the major cause of acute kidney failure in children.

We have crystallized and solved the structure of the B subunit of VT-1, a member of the Shiga toxin family. The B subunit forms a pentamer that recognizes target cell surfaces by binding to the carbohydrate component of a cell-surface glycolipid, globotriaosylceramide (Gb_3). Since this interaction largely determines which cells the toxin will attack, we are interested in understanding its molecular details. Comparison of the amino acid sequences of all the toxins in this family has enabled us to predict that a surface cleft between B-subunit monomers will prove responsible for binding. Crystallographic binding studies should allow us to test this prediction.

The most surprising result is an unexpected structural similarity to the B subunit of members of the cholera toxin family. This is unexpected because the associated A subunits of the two toxin families are completely unrelated, the B

subunits are very different in size, and the degree of amino acid sequence similarity is, if anything, less than what one would expect for two random sequences. Nonetheless, the structural similarity implies some distant evolutionary relationship between these families.

Pseudomonas aeruginosa Pilin

Pseudomonas aeruginosa is an opportunistic pathogen that infects burn victims and immunocompromised patients. It is also one of the major pathogens infecting the lungs of cystic fibrosis patients. Filaments on its surface, termed pili, attach to epithelial cell surfaces, promoting colonization. Pili are formed from a helical array of identical pilin subunits. We have crystallized pilin from this organism (in collaboration with William Paranchych, University of Alberta) and are attempting to improve the quality of the crystals to allow us to determine its structure. The three-dimensional structure would help us understand the details of cell-surface binding, which could be used to devise strategies to interfere with colonization.

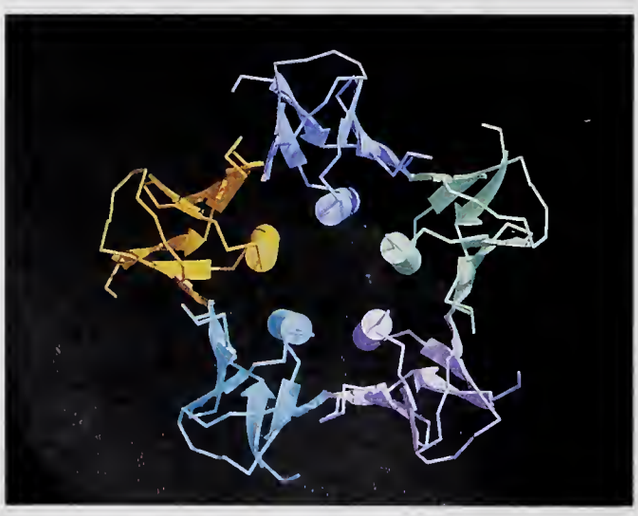
Computer-aided Drug Design

Most of the drugs in use today were discovered by trial and error, and many of these have unknown mechanisms of action. Those with known mechanisms usually act by binding specifically to a drug receptor, often a protein molecule. Since crystallography allows us to examine the structures of receptors and the details of their interactions with drugs, it should help in improving existing drugs or even in inventing new ones. Progress has been made through crystallography in the former objective, but designing new drugs from scratch is still extremely difficult. We are trying to address this problem with the aid of computers.

To design a new drug, one must first choose an appropriate receptor or target. In the design of an antibacterial drug, the key factor is selective toxicity: the drug should poison the pathogen but not the patient. One might choose, as a drug target, an essential enzyme in a biochemical pathway unique to the bacterium, because that would minimize the chances of side reactions with host enzymes. In our work on drug design, we are assuming that a good potential drug target has been chosen and its crystal structure determined. The problem, then, is how to exploit the structural information. There are probably billions of compounds that might be used as drugs; determining which of these might bind to, and interfere with, the target protein is far from trivial.

We have chosen a "divide and conquer" approach to drug design to reduce the magnitude of this problem. The vast set of possible compounds is made up of various combinations of a much smaller set of molecular fragments. We propose to design drugs by using a computer to dock members of a library of fragments to the region of the desired binding site, then to combine docked fragments to form chemically sensible molecules. We have tested the feasibility of fragment docking, and the results are sufficiently promising that we will go on to test the feasibility of the next step. Fragment docking requires the calculation of binding energies, which as yet can only be approximate and requires a great deal of computer time. But we believe that this method can provide a useful tool for drug design in the not-too-distant future, in view of the acceleration in the speed of computers and the improvements that are continually being made in the understanding of molecular interactions.

Our work is also supported by the Medical Research Council of Canada and the Alberta Heritage Foundation for Medical Research.



A schematic illustration of the crystal structure of the B (binding) subunit of verotoxin-1, viewed from the top of the pentamer. Arrows indicate strands of β-sheet, and cylinders indicate α-helices.

From Stein, P.E., Boodhoo, A., Tyrrell, G.J., Brunton, J.L., and Read, R.J. 1992. Nature *355:748–750. Copyright © 1992 Macmillan Magazines Limited.*

Representations and Transformations of Tactile Signals in Somatic and Frontal Motor Cortices

Ranulfo Romo, M.D., Ph.D.—International Research Scholar

Dr. Romo is Professor of Neuroscience at the Institute of Cellular Physiology, National Autonomous University of Mexico, Mexico City. He received his M.D. degree from the National University of Mexico and a Ph.D. degree in neuroscience from the University of Paris, France. His postdoctoral work was done with Jacques Glowinski at the College of France in Paris, Wolfram Schultz at the University of Fribourg, Switzerland, and Vernon B. Mountcastle at the Johns Hopkins University in Baltimore. Dr. Romo has received the Demuth Prize of the Swiss Medical Research Foundation and a Guggenheim Memorial Foundation fellowship.

PRIMATES have access to events occurring in the external world through specialized sensory systems. The events are first transduced by the sensory receptors and encoded and transmitted to the central nervous system by the primary afferent fibers. These messages are orderly distributed and processed in brain centers, where the external events are represented and, under some conditions, lead to sensation, perception, memory, and purposeful motor acts. These phenomena can only be studied in highly evolved brains. Our laboratory is investigating the representation of sensory signals in the brain and the mechanisms by which the motor centers process them in order to guide behavior.

The somatic sensory system of subhuman primates appears to be an appropriate model for approaching the question of how tactile signals are represented in the cerebral cortex, since the hands and relevant brain structures are much like those of the human. Similar sensory performance in somatesthetic tasks has been observed in both. Moreover, the exploratory hand movements of both primates have similar characteristics, reflecting the fact that their somatic and motor systems are similarly linked. These parallels set the base for studying the dynamic neural operations of the sensory-motor interface—in other words, the way sensory representations guide motor behavior.

We have selected an experimental paradigm in which monkeys discriminate among sensory stimuli delivered to the skin of their hands as we record the associated cortical activity. This allows us 1) to define the relevant stimuli among which monkeys can discriminate, 2) to follow the transformation of the initial cortical display in the distributed cortical system separating it from those cortical areas that drive the differential motor responses required for successful execution of the task, and 3) to study the details of intracolumnar operations in the sensory-association areas of the parietal lobe. I will refer to the first point.

Experiments are in progress to define how the direction and speed of a probe moving across glabrous skin of behaving monkeys is represented in the activity of somatosensory cortex neurons (areas 3b and 1). For this, we have designed and constructed a Cartesian robot that allows mechanical stimuli to be presented to the skin of an awake primate's hand at specified traverse distance, speeds, and directions. We have quantitatively studied many neurons in areas 3b and 1 of two alert monkeys performing a behavioral task unrelated to the tactile stimuli. The receptive fields were scanned with a probe moving at different speeds in eight directions at preselected levels of force exerted by the probe in the skin.

The first objective in the analysis of the responding neurons was to reconstruct the initial representations of the peripheral events in areas 3b and 1 and to identify the possible transformations occurring in these stages of the somatic processing system. The preliminary analysis indicated that it was possible to quantify the representation of the physical aspects of the stimuli in the discharges evoked: velocity and position (kinematics) and force (dynamics). The activity of a large percentage of the recorded cells in areas 3b and 1 varied with the speed of the stimulus and displayed directional preference. This finding suggests that certain sets of neurons of the primary somatic areas respond and make a neural replica of the mechanical stimuli and that there are already some transformations (preference for direction) at this level. Moreover, the effects of these two variables (speed and direction) were modulated by the force exerted by the stimulus on the skin.

What is interesting in our data is the fact that directionality can already be detected at the very beginning of the cortical somatic processing system. This finding may explain the presence of some nonisomorphic images of complex tactile signals in cortical areas 3b and 1.

We are presently studying the representation of these signals in areas 2, 5, and 7b, applying the same experimental protocol. We are also imple-

menting a method for recording the simultaneous activity of up to eight neurons in order to study the input-output operations of columns of areas 3b and 1 as animals detect and categorize between different directions and velocities of the stimuli moving across the skin.

Anterior-Posterior Patterning in the Early Mammalian Embryo

Janet Rossant, Ph.D.—International Research Scholar

Dr. Rossant is Professor of Molecular and Medical Genetics at the University of Toronto and Senior Scientist at the Samuel Lunenfeld Research Institute of Mount Sinai Hospital, Toronto. She received her undergraduate training in zoology at Oxford University and her Ph.D. degree from Cambridge University. Her postdoctoral training was in Richard Gardner's laboratory at Oxford. Before joining Mount Sinai Hospital, she was Associate Professor of Biological Sciences at Brock University, St. Catherines.

SHORTLY after implantation in the mother's uterus, the mammalian embryo is transformed from an undifferentiated group of cells with no axis of symmetry into a trilayered structure with anterior-posterior (A-P) and dorsal-ventral (D-V) polarity and the beginnings of segmentation. One of the main challenges in mammalian development is to understand the cellular processes that underlie these events and how they are genetically controlled.

Studies in vertebrates other than the mouse have implicated inductive interactions between tissue layers as important in both determining new tissue types and establishing regional domains along the A-P axis. We have devised an explant-recombination system that allows us to address the importance of such inductive interactions in the mouse. Using the mouse homeobox-containing *engrailed*-like (*En*) genes as markers for a specific anterior domain of the nervous system, we have shown that expression of these genes in ectoderm depends on interaction with underlying mesoderm.

Isolated anterior ectoderm from premesoderm stages of development will not initiate *En* gene expression *in vitro* but will express *En* genes after aggregation with later anterior mesoderm. Anterior mesoderm will also induce *En* expression in posterior ectoderm, which never normally expresses En proteins. Posterior mesoderm is incapable of En induction when combined with either early ectoderm or later posterior ectoderm.

These experiments and others in progress indicate that there is regionalization in the capacity of mesoderm to induce specific neural structures. Furthermore, preliminary evidence suggests that the basic patterning of the forebrain, midbrain, and anterior hindbrain structures is laid down by mesoderm induction at the early neural plate stage and that the later interactions that lead to the localized expression of other anterior genes may be confined to the neuroectoderm.

We are continuing these kinds of studies on both mesoderm and neural induction in the mouse, making use of the increasing number of early marker genes for the processes of A-P patterning. We expect these studies to define in more detail the temporal and spatial parameters of inductive interactions. The data will help in the search for the underlying molecular basis of patterning in the embryo.

One factor thought to play a role in A-P patterning is retinoic acid (RA). There is considerable circumstantial evidence to implicate RA in helping to establish the boundaries of specific A-P domains in the developing embryo. We have provided more such evidence by showing that transgenic mice, carrying an RA-responsive element upstream of a neutral promoter-*lacZ* construct, express the bacterial *lacZ* gene in a specific posterior domain of the embryo.

The boundary of this expression domain coincides, at the early neural plate stage, with the anterior boundary of the *Hox* gene, *Hox-2.9*. As development proceeds, the boundary of expression of the transgene recedes in concert with the establishment of more-posterior *Hox* gene boundaries in the hindbrain. *Hox* genes are thought to be involved in A-P patterning, especially in defining the identity of the segmental hindbrain rhombomeres. The coincidence of *Hox* gene boundaries and the receding RA-responsive transgene boundary suggests a possible involvement of RA in establishing *Hox* gene expression domains.

We have shown that anterior members of the *Hox-2* gene cluster respond within four hours to exogenous RA *in vivo*. This response involves an anterior shift in their expression boundaries, and the altered expression domain persists in later embryos and can be correlated with the specific teratogenic effects of RA in the hindbrain. Further experiments are planned to confirm that endogenous RA is important for *Hox* gene patterning.

Our knowledge of the genetic control of patterning in the gastrulating mouse embryo is still very limited. If one could screen rapidly through large numbers of genes for their pattern of expression at gastrulation, one could hope to identify overlapping expression patterns that indicate fundamental developmental domains. In collabo-

ration with Alexandra Joyner (HHMI, International Research Scholar), we are using gene-trap vectors inserted into embryonic stem (ES) cells as one strategy for identifying expression domains.

In these vectors the bacterial *lacZ* gene lacks a promoter but has a splice acceptor sequence upstream, such that *lacZ* can be activated as a fusion transcript when the vector inserts in the intron of an active host gene. Many *lacZ*-expressing ES clones with different integrations are being assessed in chimeras for their pattern of expression at gastrulation. We have screened more than 200 such lines to date and find about 20 percent with spatially restricted expression patterns. The screen will continue, and the most interesting lines will be cloned from the fusion transcript and taken through the germline to analyze the mutant phenotype induced by the vector integration.

Other more rapid screens for expression patterns using whole-mount *in situ* hybridization with cDNA clones are in the pilot stages. The combination of large-scale searches for new patterns of gene expression with embryonic manipulation experiments using such genes as markers should give new insights into the basis of patterning in the mouse embryo. This broad approach does not, of course, eliminate the need to study further the role of specific genes already identified as potential regulators of development, and we are also undertaking the targeted mutagenesis of a number of such genes.

Mouse embryos at day 8.5, showing expression domains of RAREhsplacZ *(top),* Hox-2.9 *(middle), and* Krox-20 *(bottom).*

Research and photograph by Ron Conlon in the laboratory of Janet Rossant.

Response of the Cerebral Cortex to Spatial Information

Jean-Pierre Roy, M.D.—International Research Scholar

Dr. Roy is Assistant Professor of Neurology and Neurosurgery at the Montreal Neurological Institute, McGill University, and Neurologist at Montreal Neurological Hospital, Montreal, Quebec. After receiving his M.D. degree from Laval University in Quebec City, he spent two years in Mircea Steriade's laboratory at Laval investigating the origin of rhythmic oscillations in the thalamic relay cells. He then went on to a neurology residency at the Montreal Neurological Hospital. This was followed by fellowship studies at NIH with Robert Wurtz on neuronal sensitivity to visual stimuli in the cerebral cortex.

THE main goal of my work is to understand the transformation of simple information, such as the speed and direction of motion registered by individual neurons, into information about the motion of the subject and the spatial properties of his environment. My approach is to examine the response of single cells in one area of a rhesus monkey's parietal cortex—the medial superior temporal area (MST)—to different visually presented stimuli. The neurons in the MST are interesting in that about half of them respond to motion in one direction, for example left, when the motion is in front of where the subject is looking, and respond to motion in the opposite direction, right in this case, when the motion is behind where the subject is looking. This corresponds, we think, to the apparent motion of the environment experienced during self-motion.

When a subject moves in the environment while looking at an object to the side, the objects in the foreground will move in the direction opposite to that of self-motion while those in the background will move in the same direction as the self-motion. The correlation between direction of motion and disparity in these neurons suggests that they detect these environmental motions. But more than that, it suggests that these cells are involved not so much in indicating the movement of the environment as such, but rather the movement of the subject himself. Indeed, these cells will respond if foreground moves left or if background moves right, or both. Each of these three conditions will be observed when the subject moves to the right. Hence we propose that those neurons signal motion of the subject himself.

Within one direction and one disparity response, we have evidence that speed is another property the cells record. Indeed, there appears to be a preference for lower speeds when the disparity of the motion is of a small absolute value, i.e., when it corresponds to motion close to the point of fixation, and a preference for higher speeds when the disparity is of a large absolute value, i.e., when it corresponds to motion far from the point of fixation. This last property is very interesting in that it suggests that those cells could be signaling not only the direction of the self-motion but possibly also the structure of the environment.

This requires elaboration. I am proposing that the preliminary evidence of a differential response of cells to different speeds could represent a role for them in transforming the speed information into depth (or more precisely, relative depth) information.

During the condition described above, objects in the environment that are very near or very far will move fast, while those that are close to the plane where the subject fixates will move slowly. Speed, then, contains information about the depth of objects in the environment. In order to test this hypothesis, individual planes moving independently, as were used before, are inadequate. What is needed are multiple planes moving simultaneously at different speeds. To explore whether MST neurons do respond to these stimuli representing speed and disparity gradients, we have developed stimuli with appropriate characteristics.

With those stimuli, we should be able to examine the response of neurons to different simple properties of the stimulus. Our prediction is that the cells should respond optimally to a stimulus that has the correct direction of motion of the foreground versus the background, but also a speed gradient and possibly a disparity gradient. This would suggest that those cells are indeed involved in transforming the information they receive about speed into information about depth. It would not be the ultimate answer, which could only come from lesion studies we are planning for the future.

This work is also supported by a grant from the Medical Research Council of Canada.

Molecular Studies on Neuronal Calcium Channels

Terry P. Snutch, Ph.D.—International Research Scholar

Dr. Snutch is Assistant Professor of Zoology and Neuroscience at the Biotechnology Laboratory, University of British Columbia, Vancouver. He obtained a B.Sc. degree in biochemistry from Simon Fraser University, Vancouver, and remained there to complete his Ph.D. degree for studies on the molecular genetics of the heat-shock response of Caenorhabditis elegans *with David Baillie. After further research training in the laboratory of Norman Davidson at the California Institute of Technology, Dr. Snutch joined the newly formed Biotechnology Laboratory. He recently received a fellowship in neuroscience from the Alfred P. Sloan Research Foundation and the Killam Research Prize from the University of British Columbia.*

THE entry of calcium ions (Ca^{2+}) into cells mediates a wide variety of cellular and physiological responses, including muscle contraction and hormone secretion. In the nervous system, an increase in intracellular Ca^{2+} concentration directly affects the electrical properties of neurons. Ca^{2+} entry has also been shown to have a role in regulating gene expression, modulating Ca^{2+}-dependent enzymes, and mediating nerve growth and regeneration. Furthermore, an increase in Ca^{2+} concentration at the presynaptic nerve terminal triggers the release of neurotransmitter, inhibiting or exciting postsynaptic neurons.

The rapid entry of Ca^{2+} into neurons is mediated by membrane proteins called Ca^{2+} channels. These diverse molecules respond to voltage changes across the cell membrane by opening Ca^{2+}-selective transmembrane pores. Besides the wide variety of normal physiological effects that Ca^{2+} channels mediate, they are implicated in a number of disorders, such as angina, hypertension, migraine, and certain arrhythmias. The clinical treatment of these disorders is aided by Ca^{2+} channel–blocking drugs. Our present studies utilize molecular cloning techniques to address basic questions concerning the structure, function, and expression of Ca^{2+} channels in the mammalian nervous system.

Molecular Diversity of Neuronal Ca^{2+} Channels

Four types of Ca^{2+} channel, called T, L, N, and P, have been identified, with differing electrophysiological and pharmacological properties. Among their subunits is the α_1-subunit, which both responds to voltage changes and forms the ion-conducting pore. Although the differences among various Ca^{2+} channels may be due to a number of factors, we hypothesize that distinct α_1-subunits account for most Ca^{2+} channel diversity.

Utilizing molecular cloning techniques, we have isolated cDNAs encoding four distinct classes of α_1 Ca^{2+} channel from rat brain: classes A, B, C, and D. Their amino acid sequences show that they are large proteins (2,100–2,300 amino acids) and that they share a number of conserved features with cloned sodium and potassium channels.

The class C channel found in brain is nearly identical to a Ca^{2+} channel previously found in heart and lung, while classes A, B, and D represent novel forms. Hybridization of the clones to rat genomic DNA indicates that each of the classes of brain Ca^{2+} channel is encoded by a distinct gene.

One interesting result of these studies is that within each class of α_1-subunit cloned, several varieties have been identified, and the actual number of distinct α_1-subunits is much larger than previously thought. Using molecular genetic techniques, we have demonstrated that these varieties, at least in the class C and D instances, are generated by alternative splicing. For class C Ca^{2+} channels, two distinct isoforms are expressed at different levels in various regions of the rat brain and thus may make unique contributions to distinct populations of neurons.

Differential Localization of Ca^{2+} Channel Subtypes

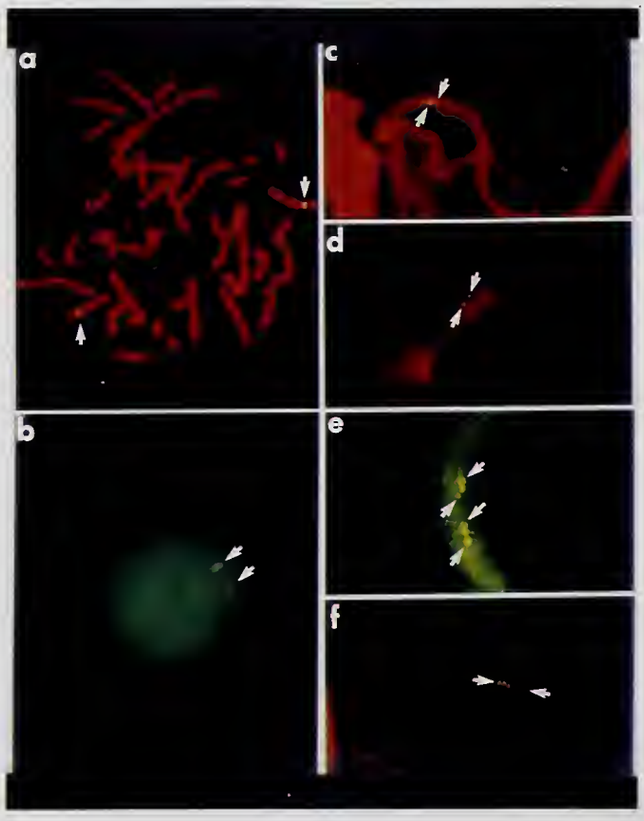
In an attempt to provide information concerning the physiological roles of the neuronal Ca^{2+} channel subtypes, we are determining their cellular and subcellular localizations. Studies localizing Ca^{2+} channel gene expression by examination of RNA levels show that the various classes of Ca^{2+} channel are differentially expressed in the brain. For example, in the cerebellum the class A channel is mostly localized to Purkinje cells, while class C channels are more widespread. In addition to their different spatial distributions in the brain, studies using antibodies generated against the cloned Ca^{2+} channel α_1-subunits show that the various Ca^{2+} channel proteins have distinct subcellular localizations on individual neurons. For example, some Ca^{2+} channel subtypes appear to be localized to dendritic regions, while others are located on both cell bodies and dendrites. We believe that the distinct cellular and

subcellular distributions of the Ca^{2+} channel subtypes reflect their unique contributions to neuronal physiology. A long-term goal of this research is to define how the individual subtypes uniquely contribute to neuronal functioning.

Functional Expression of Neuronal Ca^{2+} Channels

We are taking two approaches to determining the functional properties of the cloned Ca^{2+} channel α_1-subunits. First, full-length clones for the four main classes of rat brain α_1-subunit are being introduced into mammalian cell lines that normally do not express any Ca^{2+} channels. The electrical and pharmacological properties of the Ca^{2+} channels will then be determined with the patch-clamp technique. Second, we are using the DNA sequences derived from the cloned Ca^{2+} channels to generate small synthetic pieces of DNA (oligonucleotides) that will hybridize to mRNAs encoding these molecules.

Microinjection of oligonucleotides into cells that already express defined Ca^{2+} channels results in the oligonucleotide hybridizing to Ca^{2+} channel mRNA and inhibiting expression. By inhibiting the expression of a single type of Ca^{2+} channel in cells that normally express several types, we hope both to identify the specific type of Ca^{2+} channel blocked by the oligonucleotide and also to obtain some insight into the physiological roles that individual Ca^{2+} channel types contribute to neurons. Using this technique we have found that the rat brain class C α_1-subunit encodes a dihydropyridine-sensitive (L-type) Ca^{2+} channel. Similar studies with the other classes of brain Ca^{2+} channel are under way.



Detection of single-copy sequences in fluorescence micrographs of metaphase chromosomes, interphase nuclei, and free chromatin. a) Metaphase chromosomes from a human diploid lymphocyte culture, hybridized with four cosmid probes (cM58-3.6, CF14, cJ21, and cW10-20) together, spanning 341 kb. b) Hybridization of the same probes with interphase nucleus. Arrows indicate two sets of hybridization signals. c–f) Results of hybridization with different combinations of cosmids in the 7q31 region. c) cNH24 and cJ21; d) cM58-3.6 and cJ21; e) cM58-3.6, CF14 and cJ21 (note: two sets of hybridization signals); f) cM58-3.6, CF14, cJ21, and cW10-20.

Research and photograph by Henry Heng in the laboratory of Lap-Chee Tsui.

Cystic Fibrosis, Gene Expression in the Mammalian Lens, and Mapping of Chromosome 7

Lap-Chee Tsui, Ph.D.—International Research Scholar

Dr. Tsui is Senior Scientist and Sellers Chair of Cystic Fibrosis Research in the Department of Genetics at the Research Institute of the Hospital for Sick Children, Toronto, and Professor of Molecular and Medical Genetics at the University of Toronto. He was born in Shanghai, raised and educated in Hong Kong, and there awarded degrees from the Chinese University. His Ph.D. degree is from the University of Pittsburgh, where his thesis was on the structure and assembly of bacteriophage λ (with Roger Hendrix). After training briefly in the Biology Division of Oak Ridge National Laboratory, he joined the laboratories of Manuel Buchwald and Jack Riordan at the Hospital for Sick Children to work on cystic fibrosis. Dr. Tsui's honors include the titles of Scientist of the Medical Research Council of Canada, Fellow of the Royal Society of Canada, and Fellow of the Royal Society of London.

THE research interests of my laboratory consist of three general topics in the molecular biology of mammalian gene regulation and function.

Molecular Genetics of Cystic Fibrosis

Through classical genetic linkage analysis and various molecular cloning strategies, the gene responsible for cystic fibrosis (CF), the commonest severe autosomal recessive disorder among Caucasians, has been localized and identified. We have shown that the major CF mutation, accounting for approximately 70 percent of all mutant alleles, is a deletion of the phenylalanine residue at amino acid position 508 of the predicted polypeptide, which is named cystic fibrosis transmembrane conductance regulator (CFTR).

To investigate the basic defect in CF, we are continuing our search for the other CFTR mutations. Since the CFTR gene contains 27 exons and spans 230 kb of DNA, detecting a microscopic mutation is not straightforward. Furthermore, the lack of a convenient functional assay for CFTR makes it difficult to distinguish a truly disease-causing mutation from a benign amino acid substitution. To coordinate the detection effort, our laboratory has taken a central role in the formation of an international consortium of 90 groups of researchers from 26 countries. Through active exchange of gene sequences and mutation data before submission for publication, the consortium has already identified more than 150 apparent CF mutations and 30 sequence variations. The collective data from two years of operation also show that most of the remaining 30 percent of mutant alleles are individually rare and highly heterogeneous among different populations.

Despite our intensive DNA sequencing effort, covering each of the exons plus their flanking regions, we have not been able to identify the mutations for 6 of the 94 CF chromosomes examined. These alleles probably harbor different mutations located in regions that affect transcription or RNA processing. Nevertheless, we have developed a highly informative marker for use in genetic counseling of CF families. The observed heterozygosity for this marker, a dinucleotide repeat polymorphism in one of the introns of the CFTR gene, is over 95 percent. In a case of prenatal diagnosis, we applied the marker successfully after all other available tests failed.

The common CF symptoms include chronic obstructive lung disease, pancreatic enzyme insufficiency, and elevated sweat electrolytes. Other organs and tissues, such as the hepatobiliary tree, intestines, and vas deferens, may also be involved. The varied degree of severity of the symptoms among CF patients suggests that the phenotypes are at least partly conferred by the genotypes at CFTR. Furthermore, information about genotypes and phenotypes should also provide important clues to the function of CFTR and the basic defect in the various affected organs and tissues. Based on this assumption and the use of the extensive clinical data collected at the CF clinic in our hospital, we have demonstrated a good correlation between pancreatic involvement and genotypes at the CFTR locus.

To facilitate direct biochemical and physiological analysis, we have constructed a full-length cDNA for CFTR in several expression vectors. Site-directed mutagenesis has been used to remove sequences that are toxic to the host bacterium as well as to generate mutant constructs for functional evaluation. Using DNA transfection into heterologous cell types, we showed that this cDNA could confer a cAMP-regulated chloride channel activity *de novo*, suggesting that CFTR is a chloride channel itself. Preliminary data from the analysis of constructs reproducing some of the naturally occurring mutations appear to be in good agreement with those predicted from the severity of pancreatic involvement.

In order to understand the regulation of the CFTR gene, we have performed a series of deletion and transfection studies to determine the sequence elements responsible for basal promoter

activity and tissue specificity. Our current data suggest that the basal promoter element is within 250 bp of the major transcription initiation site. There also seems to be a negative regulatory element immediately upstream of this sequence, and proper expression of CFTR *in vivo* may require additional cis-regulatory element(s) yet to be identified. A thorough understanding of the regulation of CFTR transcription may provide additional means for treatment of CF, particularly in some mild cases where an increase of CFTR synthesis may compensate for the partial defect.

Experiments are also in progress to exploit the yeast *STE6* gene as a genetic system to gain some insight into the structure and function of the first ATP-binding domain (NBF1) in CFTR. A systematic survey of second site mutations (which may rescue the primary mutation) should provide important information about the structure of NBF1 and the possibility of its application in drug design.

Lastly, in order to generate an animal model for the study of CF, we have been trying to inactivate the mouse Cftr gene via homologous recombination in embryonic stem cells. Our attempt to interrupt exon 10 has so far been unsuccessful, and our current targets are exons 1 and 13. The availability of a mutant mouse strain should greatly facilitate studies to clarify the pathophysiology of CF and improve means of treatment.

Regulation of Gene Expression in Mammalian Lens Development

Transparency of the vertebrate eye lens is at least partly conferred by the short-range ordering of water-soluble crystallin molecules in the lens fiber cells. In order to understand the role these proteins play in maintaining lens transparency, one of our interests has been to identify the mutation responsible for a dominant lens defect in a mouse strain called Elo (*Eye lens obsolescence*). We have performed genetic linkage analysis to show that the Elo mutation is closely linked to the γ-crystallin gene cluster and have excluded the first five of the six genes in the cluster as being the location of the mutation. Through subsequent sequence analysis, we have identified a frameshift mutation in the γE gene, the last member of the six-membered cluster. Although the significance of the latter observation is presently unknown, it strongly argues that γ-crystallin plays a major rather than a generally assumed passive role in lens development.

Physical Characterization of Human Chromosome 7

In order to generate a more complete set of reagents for the study of genes on human chromosome 7, a chromosome-specific yeast artificial chromosome (YAC) library has been constructed. Using a human-hamster somatic cell hybrid with a single human chromosome 7, we have isolated more than 1,000 YAC clones containing human DNA inserts averaging 475 kb. The clones are being mapped to specific chromosome regions by hybridization with previously localized DNA segments and with a somatic cell hybrid mapping panel. Over 100 clones have thus far been identified. In addition to the generation of a long-range physical map of chromosome 7, we are developing efficient techniques for detecting gene sequences based on YAC cloning.

We have also developed a novel *in situ* hybridization procedure to facilitate our physical mapping effort. Using chromatin fibers released from interphase nuclei with specific reagents (mostly known as topoisomerase II inhibitors) and an alkaline buffer, we have been able to perform fluorescent *in situ* hybridization to order DNA segments less than 20 kb apart. Immediate application of this technique (named free chromatin mapping) should allow us to estimate physical distance between any given DNA segments and to study complex sequence arrangement, such as that in the centromeric regions.

Index