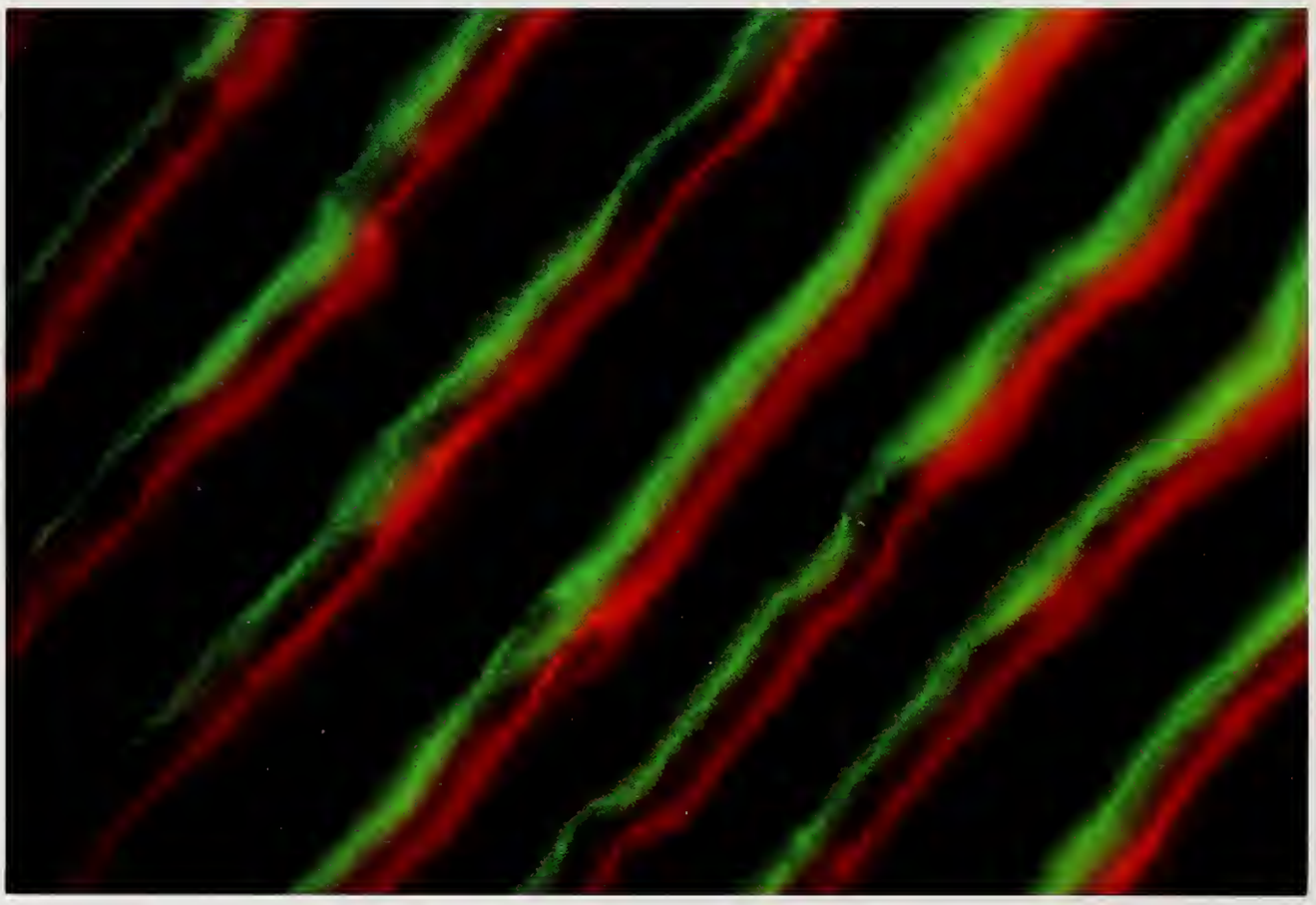
Howard Hughes Medical Institute

Research in Progress 1991

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
Executive Offices
6701 Rockledge Drive
Bethesda, Maryland 20817
(301) 571–0200



Visualization of phosphorylation of tyrosine residues in the nicotinic acetylcholine receptor at cholinergic synapses in the electric organ (electroplax) of the electric ray Torpedo. *The electroplax were immunofluorescently labeled with rhodamine-coupled α-bungarotoxin to localize the receptor (red) and with antibodies to phosphotyrosine to localize phosphotyrosine residues (green). The two fluorescent signals are superimposable; however, they have been artificially separated to allow a side-by-side comparison of the labeling patterns. Tyrosine phosphorylation of the nicotinic receptor in the postsynaptic cell is regulated by the innervating neuron and plays a role in regulating the sensitivity of the receptor to its neurotransmitter and in clustering the receptor beneath the nerve terminal.*

Research of Richard L. Huganir.

Dr. W. Maxwell Cowan
Vice President and Chief Scientific Officer
Howard Hughes Medical Institute
Printed in the United States of America

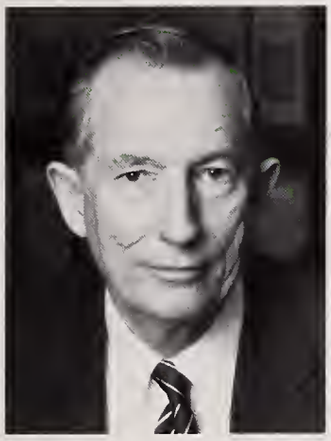
Howard Hughes Medical Institute

m

Research in Progress 1991

Contents

Foreword



THE Howard Hughes Medical Institute (HHMI) is pleased to present the sixth volume in the series *Research in Progress*. From its origin in 1986 as a means of providing a brief but informative overview of the research being carried out by the Institute's investigators, this publication has become a valuable source of information for not only the entire HHMI community but also many scientists and nonscientists who are interested in current progress in biomedical research in general. One measure of the usefulness of these reports is that last year approximately 2,000 copies were distributed, many to individuals requesting it for the first time.

New readers may be interested in other HHMI publications, some of which describe the programs of the Institute or deal with general topics in scientific research and education. The *Annual Scientific Report* is the official archival record of the Institute as a Medical Research Organization (MRO). It contains a description of the research by each HHMI investigator, including a bibliography, and highlights other activities of the Institute. A general *Annual Report of the Howard Hughes Medical Institute* describes for a lay audience the various HHMI programs and gives a summary of Institute financial data.

A series of reports for a general audience on subjects of current scientific interest is also published. The first report, produced last year and entitled *Finding the Critical Shapes*, describes exciting advances in structural biology and how the determination of the structure of biologically important molecules and the correlation of structure with function are leading to understanding of life processes and new approaches to the prevention and treatment of diseases. This report was widely acclaimed by a broad spectrum of readers, especially teachers who found it useful in their classrooms. Over 33,000 copies have been distributed. A second report, on genetics and human diseases, entitled *Blazing a Genetic Trail*, has recently been released; it is clear that it also will be highly successful and in great demand, as the first printing of 45,000 copies has almost been exhausted in a few weeks in response to many requests, including again teachers all across the country.

HHMI has a large grants program that complements its research activities as an MRO. It is anticipated that this year grant awards will total approximately $47 million. The grants program supports education in the biomedical and related sciences at the precollege, college undergraduate, graduate, and postdoctoral levels. These initiatives are described in detail in an annual publication entitled *Grants for Science Education*. In 1990 a report on *Support for Science Education and Research*, summarizing the level of funding by foundations and voluntary health associations, was also published.

In addition to providing an up-to-date report on the rapid progress in biomedical research as illustrated by the work being done in HHMI laboratories, this issue of *Research in Progress* reflects the continued expansion of the Institute's activities. As of July 1, 1991, there were 223 HHMI investigators based at 52 institutions across the nation. Their research is the core of the Institute's activities, and we invite you to share, in these pages, in the excitement of discovery and the advances in the knowledge of biological processes and disease mechanisms that hold great promise for 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
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
and 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

William T. Quillen, Esq.
Vice President and General Counsel

Robert C. White
Vice President and Chief Financial Officer

Principal Staff Members

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

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

Patricia J. Hoben, Ph.D.
Grants Program Officer

John A. Jones
Director of Computer Services

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

Laura A. Kumin, 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—Fixed Income

Dianne R. Sagner, Esq.
Associate General Counsel

Mark W. Smith
Controller

Claire H. Winestock, Ph.D.
Senior Research Program Administrator

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

John E. Dowling, Ph.D.
Maria Moors Cabot Professor of Natural Science
Department of Cellular and Developmental Biology
Harvard University

Gerald D. Fischbach, M.D.
Nathan Marsh Pusey Professor and Chair
Department of Neurobiology
Harvard Medical School

Hugh O. McDevitt, M.D.
Burt and Marion Avery Professor of Immunology
Professor of Microbiology, Immunology, and Medicine
Stanford University School of Medicine

Frederic M. Richards, Ph.D.
Sterling Professor 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

William S. Sly, M.D.
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

Thomas A. Waldmann, M.D.
Chief, Metabolism Branch
National Cancer Institute
National Institutes of Health

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
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

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

Immunology

Joseph M. Davie, M.D.
President
G.D. Searle and Company

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

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

John D. Stobo, M.D.
William Osler Professor of Medicine and Chairman
Department of Medicine
The Johns Hopkins University School of Medicine

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

Neuroscience

Floyd E. Bloom, M.D.
Chairman, Department of Neuropharmacology
Member, Scripps Clinic and Research Foundation

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

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

Melvin I. Simon, Ph.D.
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, Digestive, and Kidney Diseases
National Institutes of Health

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

Dinshaw Patel, Ph.D.
Professor of Biochemistry and Molecular Biophysics
Columbia University College of Physicians and Surgeons

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

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
Illinois	Northwestern University and associated hospitals, Evanston The University of Chicago and The University of Chicago Hospitals
Indiana	Indiana University, Bloomington, and associated hospitals
Iowa	University of Iowa and associated hospitals, Iowa City
Maryland	The Carnegie Institution of Washington and The Johns Hopkins Hospital, Baltimore The Johns Hopkins University and Hospital
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 and 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 1991. 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. In this 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 *zeppelin* 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.

As in previous years, the greater part of the volume consists of a series of short reports by the investigators associated with the Institute during the past year. The fact that this volume is appreciably larger than its predecessor simply reflects the increased number of investigators whose work is represented, since each is limited to no more than two printed pages. The reports submitted by each investigator have been collated and edited by Dr. Claire H. Winestock, Senior Research Program Administrator; Elizabeth Cowley, Copy Editor; William T. Carrigan, Editor/Writer; Gail Markley, Manager of Publications; and Kimberly A. Cornejo, Permissions Editor. 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 five years ago to reflect more accurately its principal theme—the biology of individual 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 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 influ-

ences, 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 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

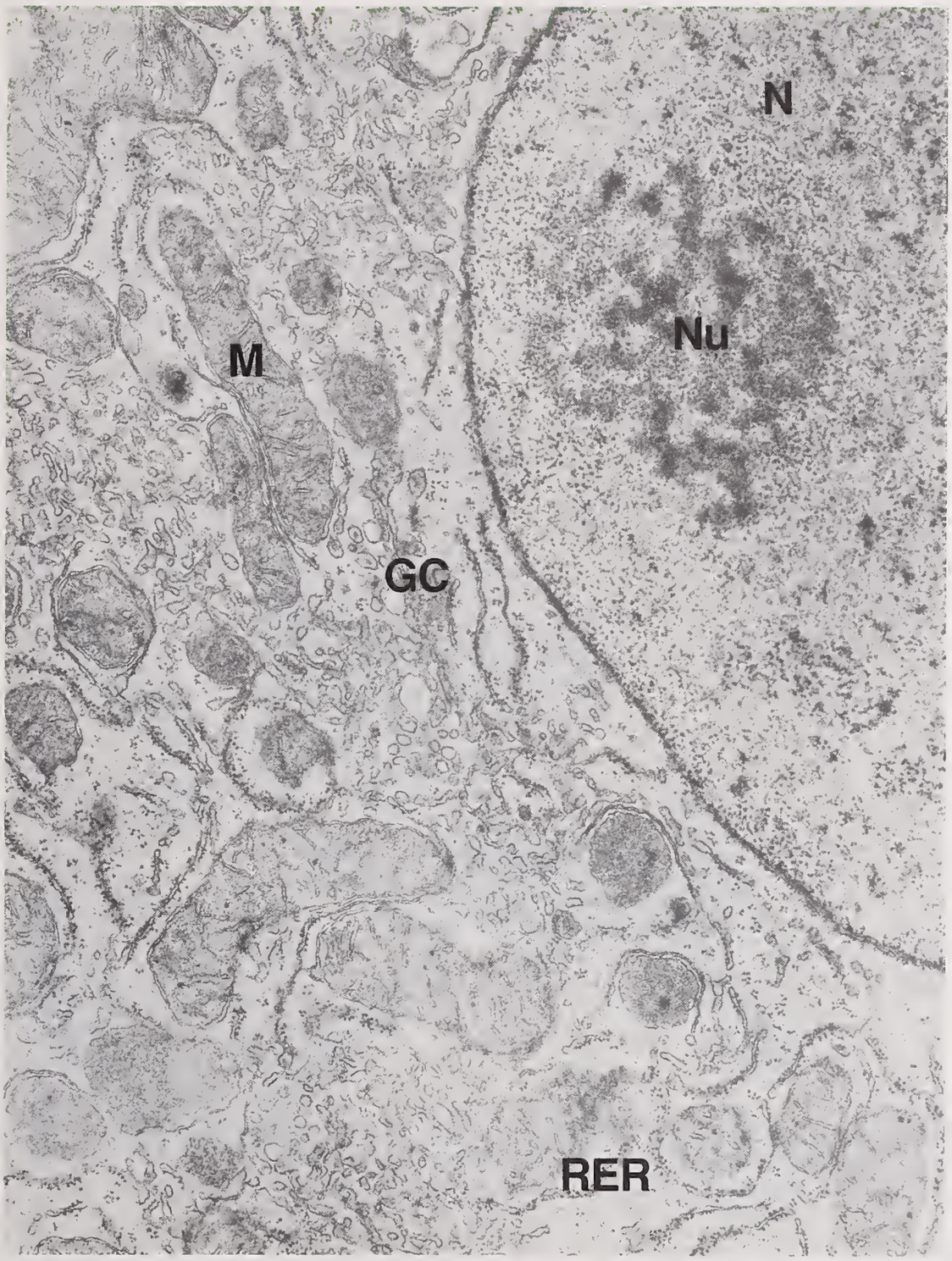


Figure 1. A small section of a typical mammalian cell as seen in the electron microscope. Part of the nucleus (N) with its surrounding membrane and nucleolus (Nu) are shown in the upper right; the cytoplasm that occupies the rest of the electron micrograph contains several different organelles, including 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.

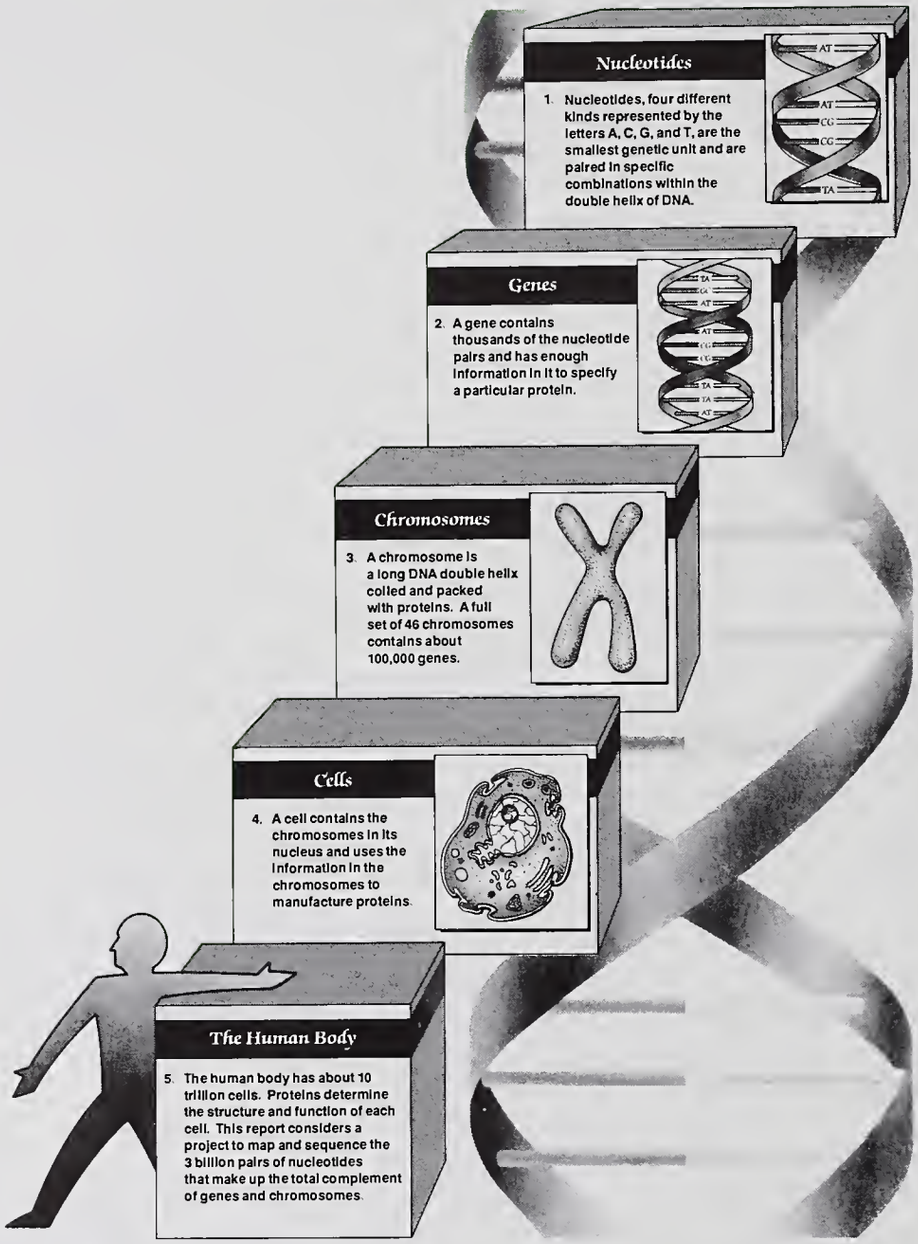


Figure 2. From genes and chromosomes to cells, organs, and entire bodies.
Adapted from an illustration by Warren Isensee for The Chronicle of Higher Education, *September 3, 1986, with permission from the publisher.*

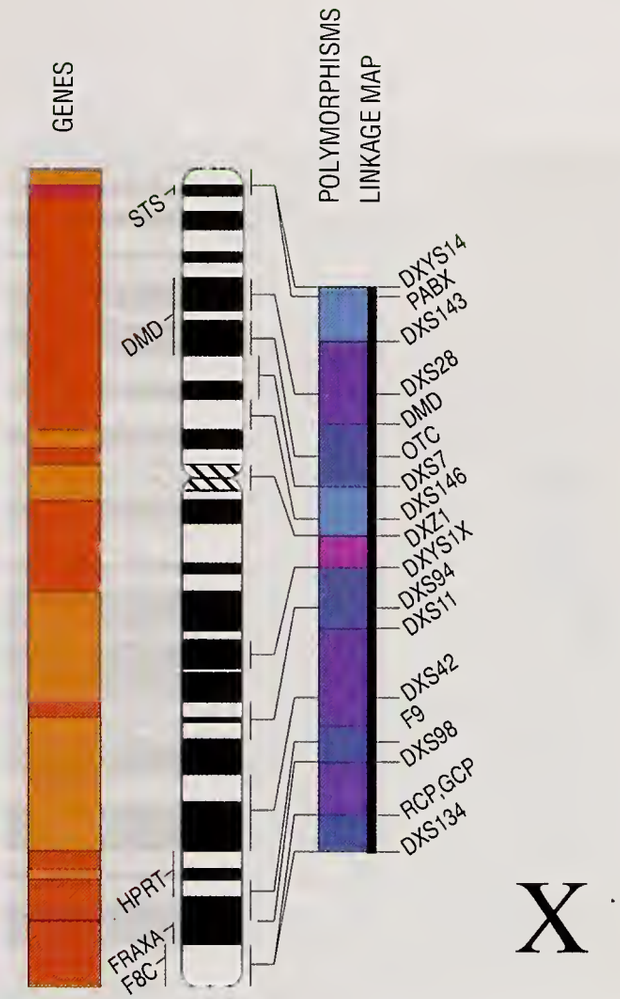


Figure 3. A schematic representation of the X chromosome showing the location of the characteristic banding patterns and the distribution of several of the genes that have been localized to this chromosome.

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.*

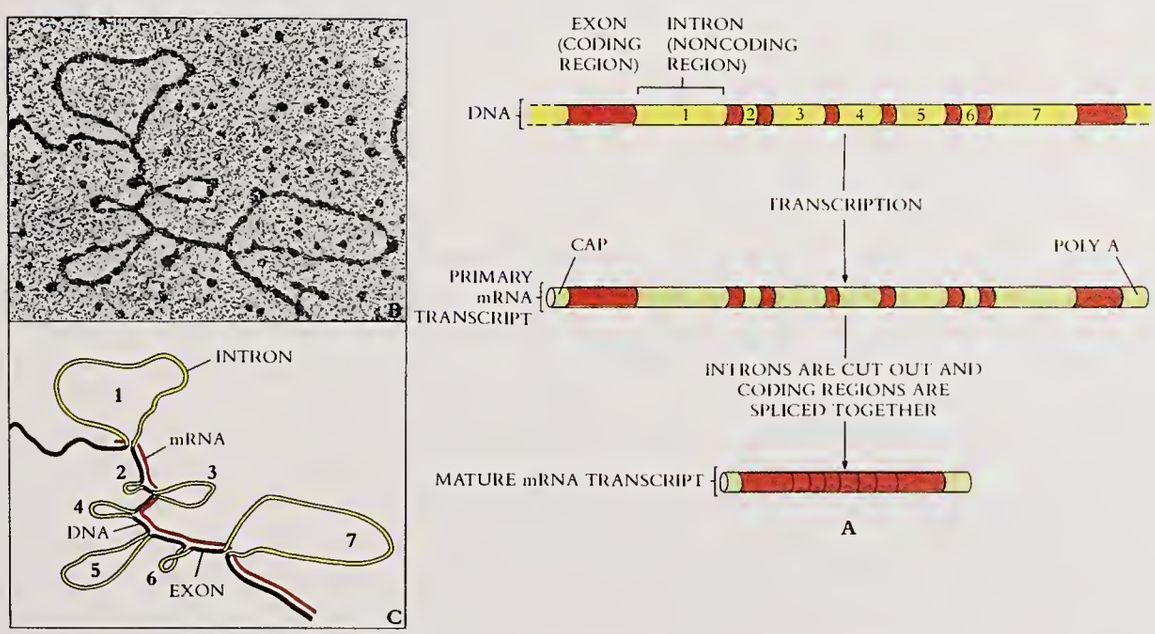


Figure 4. The primary RNA transcript of a gene is very much larger than the message that is actually translated to form a protein. A variable number of noncoding sequences called introns are edited out, and the remaining exons are spliced together by a complex mechanism shown in the drawings to the left.

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.*

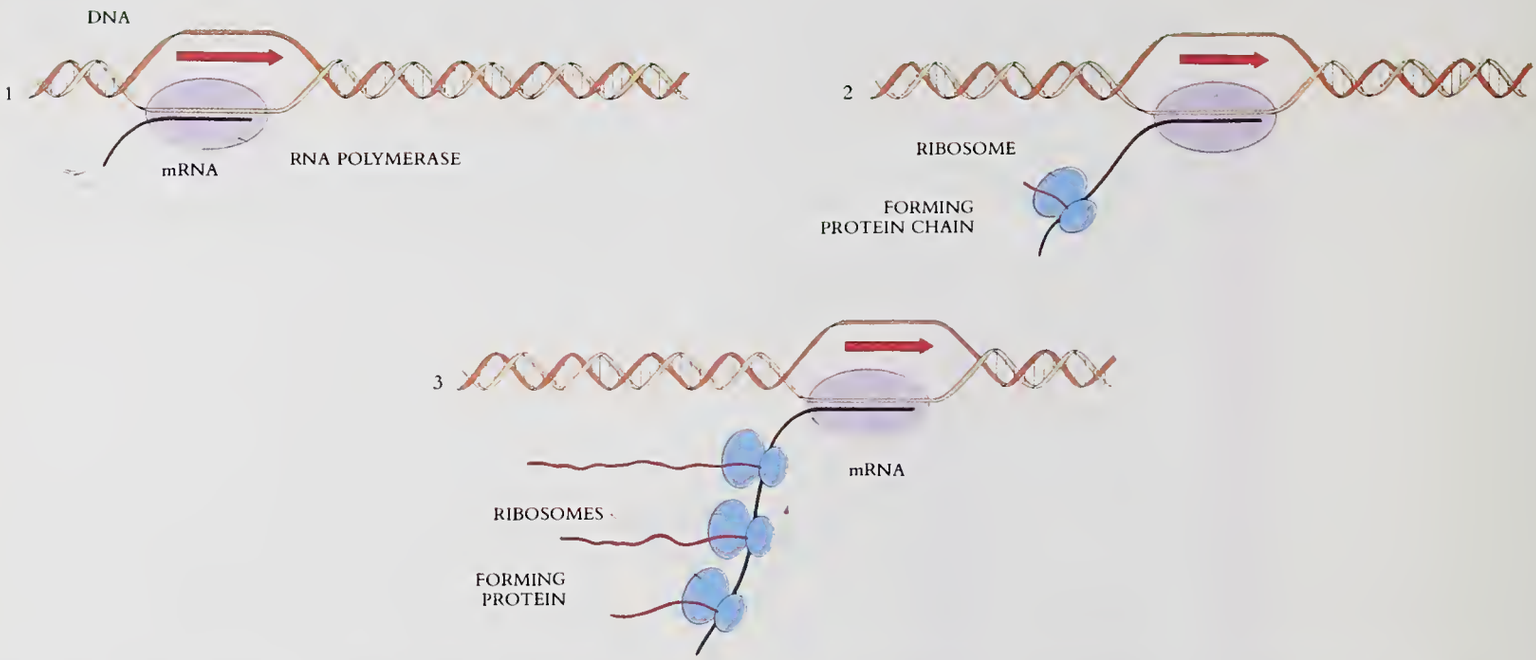


Figure 5. Transcription of a gene by the enzyme RNA polymerase and the subsequent translation of this messenger RNA (mRNA) by ribosomes to form a protein.

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

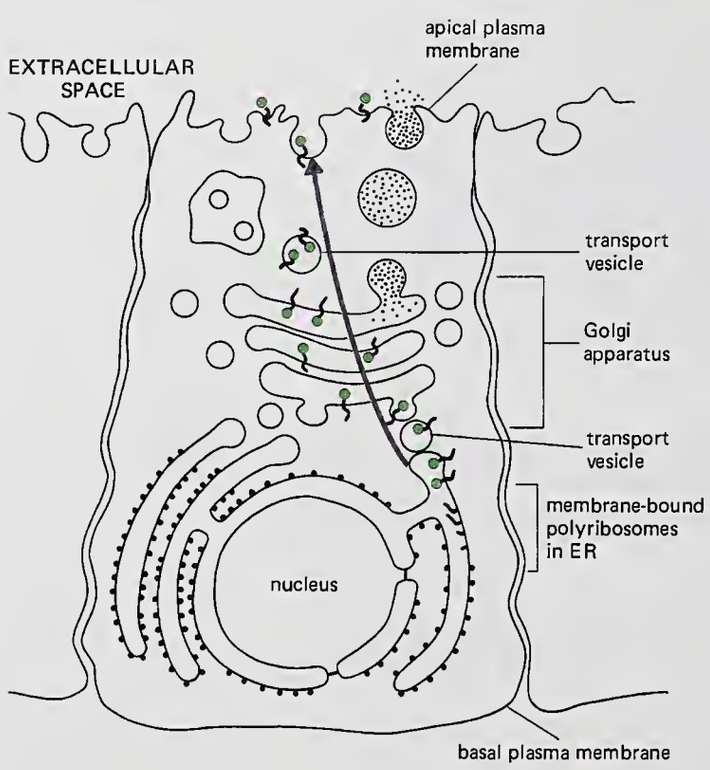


Figure 6. The synthesis and intracellular trafficking of proteins from the endoplasmic reticulum through the Golgi apparatus to the cell membrane.

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.*

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 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 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 contin-

ually 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 until death 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 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

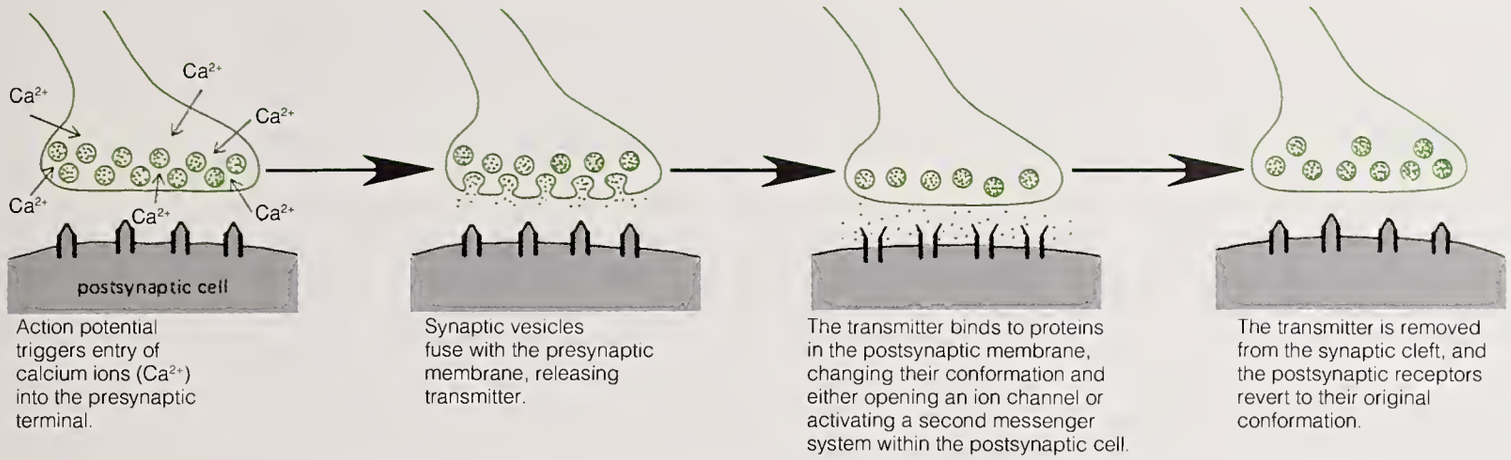


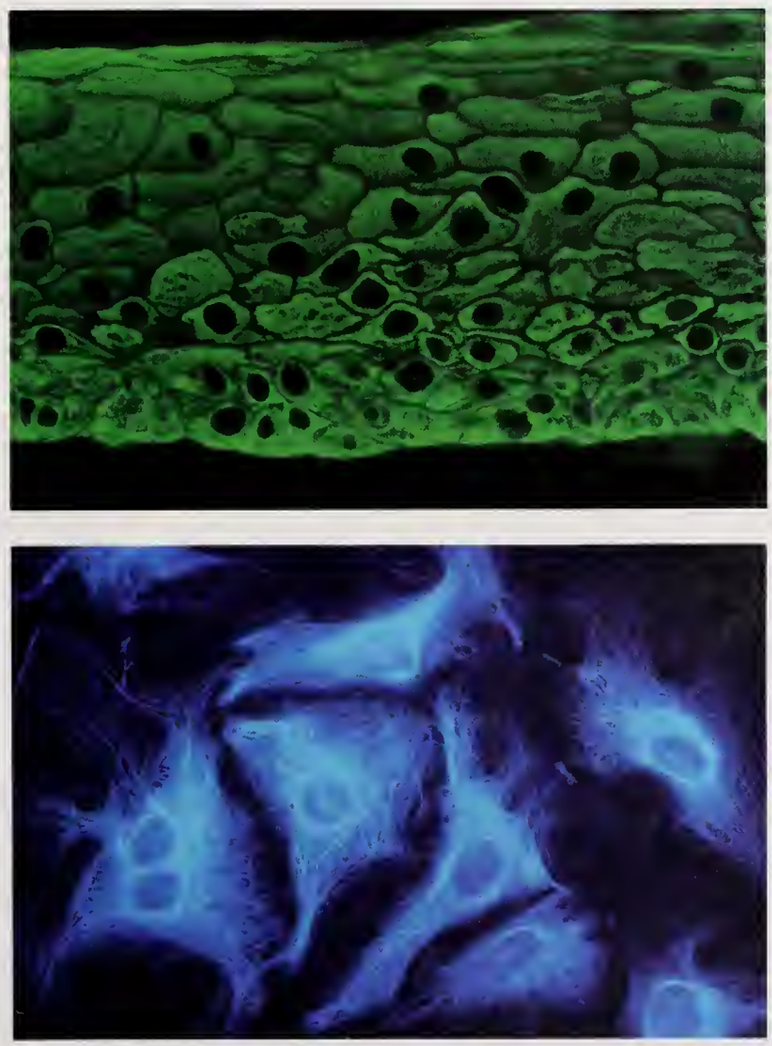
Figure 7. A summary of the essential events at a chemical synapse following the arrival of an action potential in the axon terminal.

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.



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.
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.
Carroll, Sean B., Ph.D.
Chin, William W., M.D.
Crabtree, Gerald R., M.D.
Cunningham, James M., M.D.
Davis, Roger J., M.D.
Dreyfuss, Gideon, Ph.D.
Ellis, Leland, Ph.D.
Esmon, Charles T., Ph.D.
Exton, John H., M.D., Ph.D.
Fuchs, Elaine V., 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.
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.
Larsen, P. Reed, M.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, Roeland, 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.
Spradling, Allan C., Ph.D.
Steiner, Donald F., M.D.
Watterson, D. Martin, Ph.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.

[*] *This investigator was appointed after manuscripts were submitted for 1991. 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 is by far 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 *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, and TPA (tissue 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

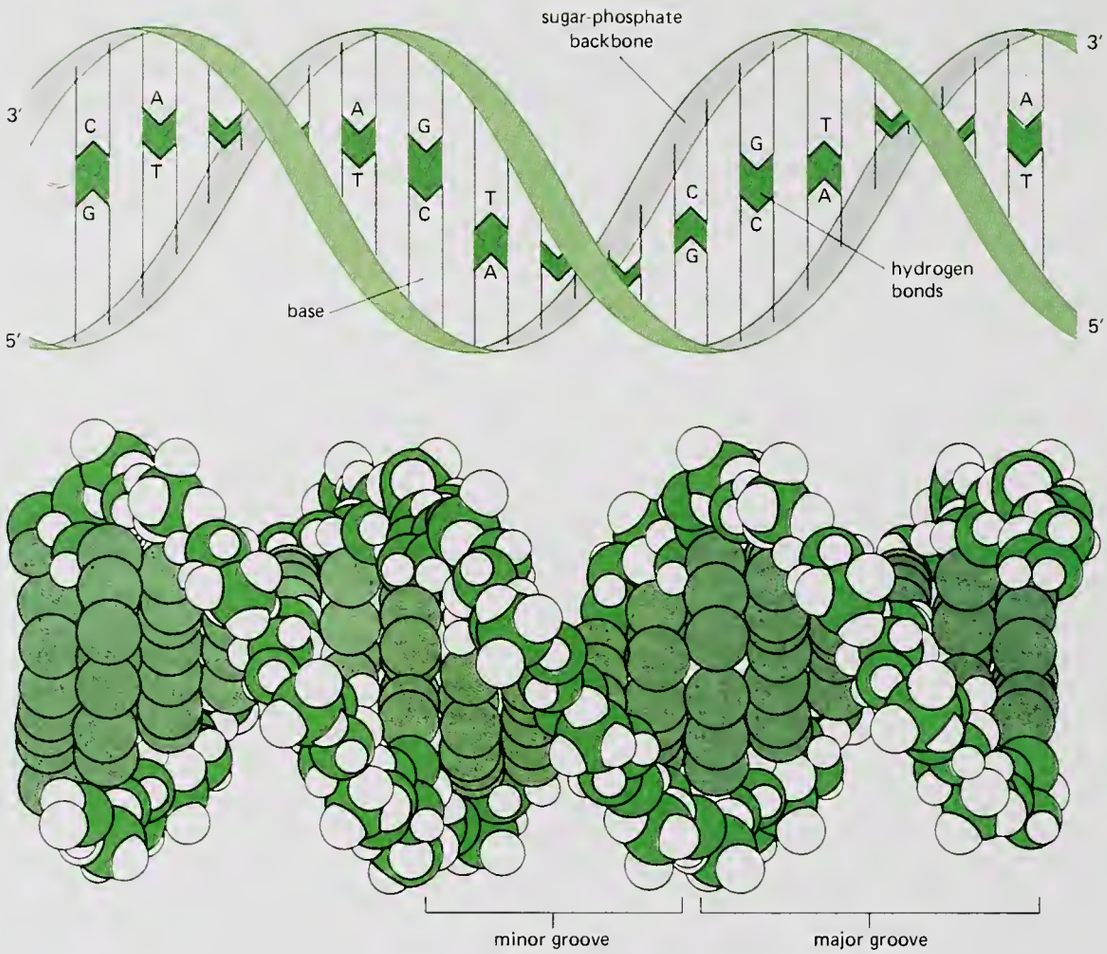


Figure 9. The famous DNA double helix.

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.*

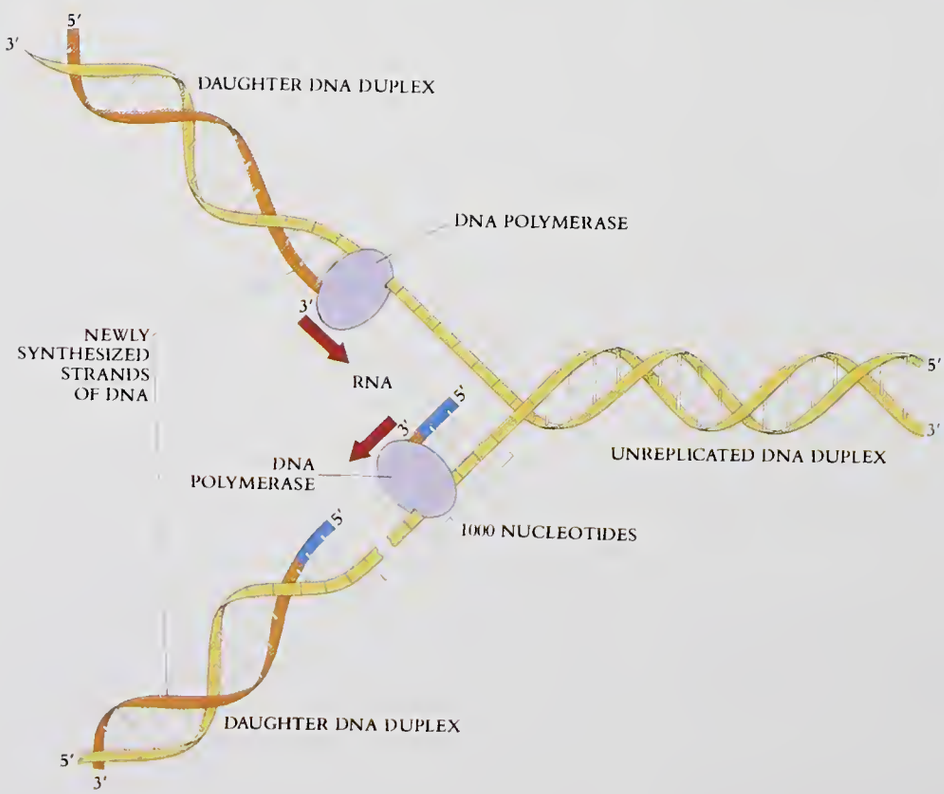


Figure 10. An illustration of how DNA is replicated so as to form two new strands that are exactly complementary to the original DNA sequences from which they derive their sequences.

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

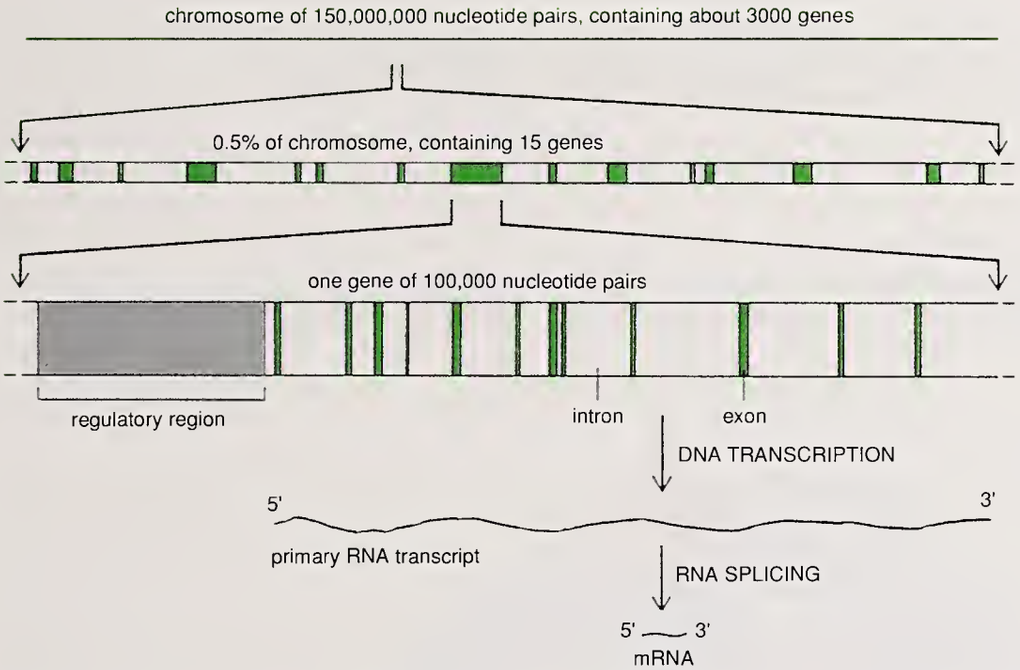
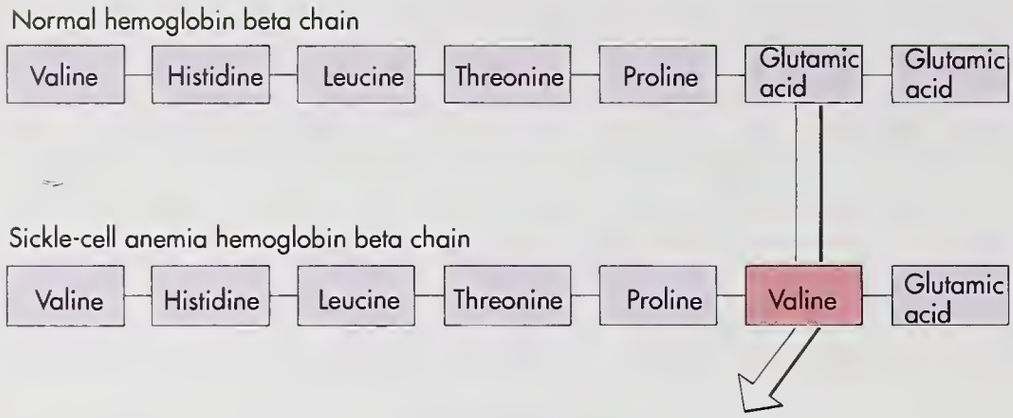


Figure 11. The organization of genes on a typical vertebrate chromosome. Proteins that bind to the DNA in regulatory regions determine whether a gene is transcribed; although often located on the 5′ side of a gene, as shown here, regulatory regions can also be located in introns, in exons, or on the 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.*



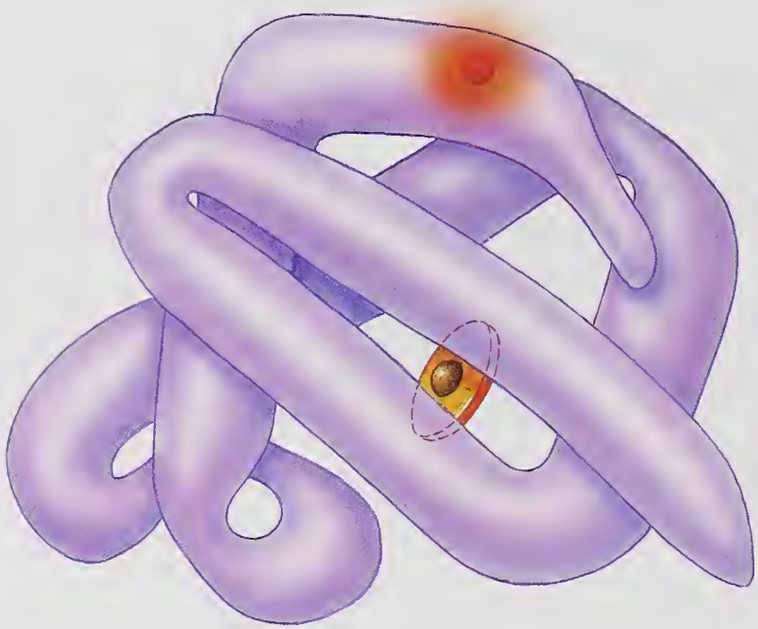


Figure 12. A drawing of the hemoglobin molecule and the single-amino acid change in the β-chain of the molecule that causes sickle cell anemia.

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

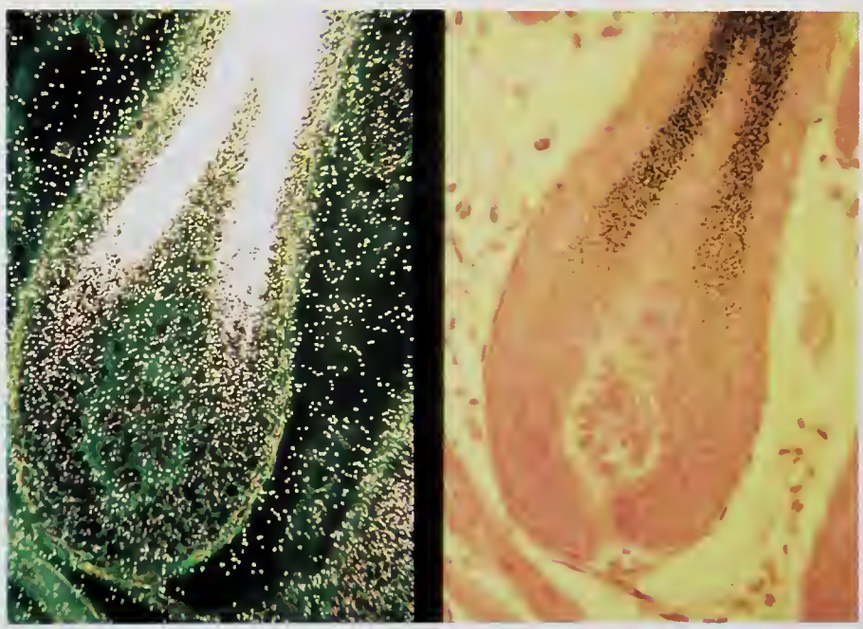


Figure 13. We can determine where particular genes are being expressed by using a radioactively labeled probe that selectively binds to the RNA transcript of the gene and examining the distribution of the radioactivity in a tissue autograph. This technique, known as in vitro *hybridization, was used 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.

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 a specific 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, 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 ac-

count, 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 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 15 or 20 years.

While the new genetics has opened the door to an understanding of an important range of fundamental problems such as the mechanism of chromosome 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.
Feinberg, Andrew P., M.D., M.P.H.
Flavell, Richard A., Ph.D.
Francke, Uta, M.D.
Friedman, Jeffrey M., M.D., Ph.D.
Geliebter, Jan, Ph.D.
George, Donna L., Ph.D.
Gesteland, Raymond F., Ph.D.
Ginsburg, David, M.D.
Gitschier, Jane M., Ph.D.
Hazelrigg, Tulle I., Ph.D.
Henikoff, Steven, Ph.D.
Horwich, Arthur, 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.
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.
Shapiro, Larry J., M.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.
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.

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**, like antibodies, 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 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 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

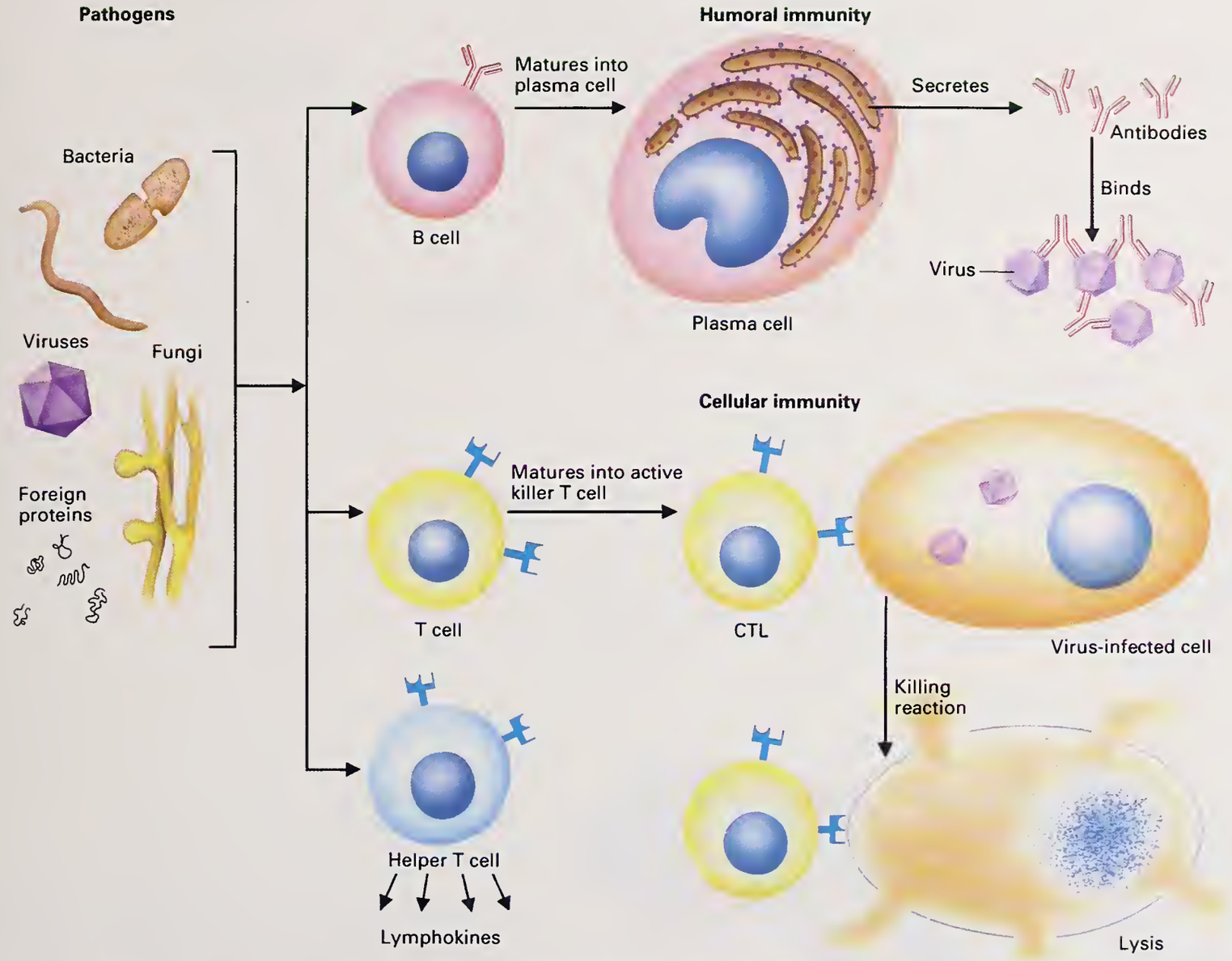


Figure 14. When a pathogen invades the body, the immune system responds with three types of reaction. The cells of the humoral immune system (B cells) secrete antibodies that can bind to the pathogen. Cells of the cellular system (T cells) carry out two major types of functions. One type of T cell (the 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.*

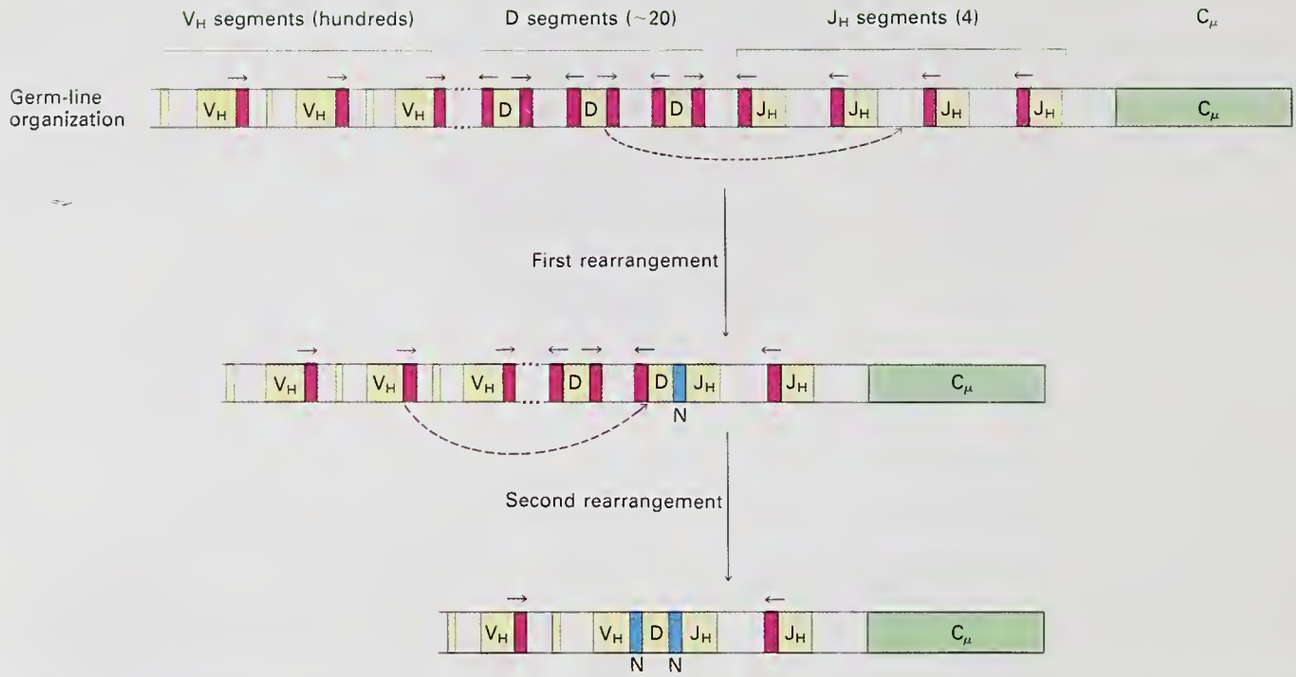


Figure 15. The specificity of antibodies and the receptors on T cells is brought about by a series of complex rearrangements of various gene segments like those shown in this figure.

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.*

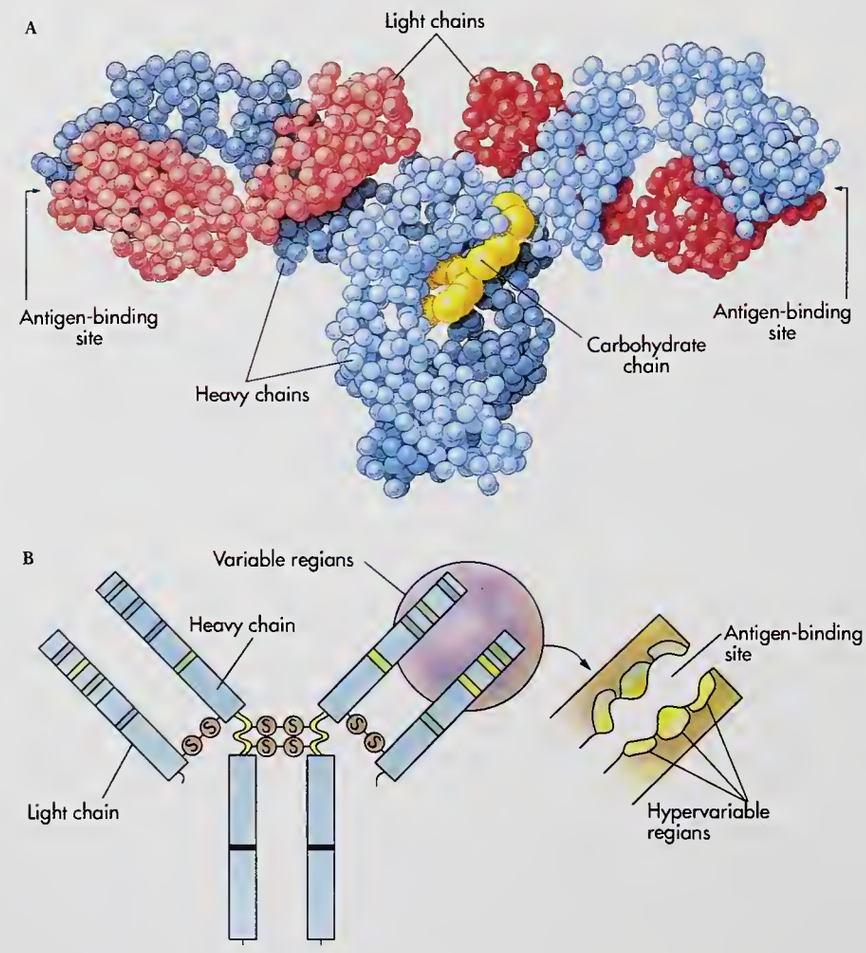


Figure 16. A: A molecular model of an antibody molecule. Each amino acid is represented by a small sphere. The heavy chains are colored blue, the light chains red. The four chains wind about one another to form a Y shape, with two identical antigen-binding sites at the arms of the Y and a tail region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

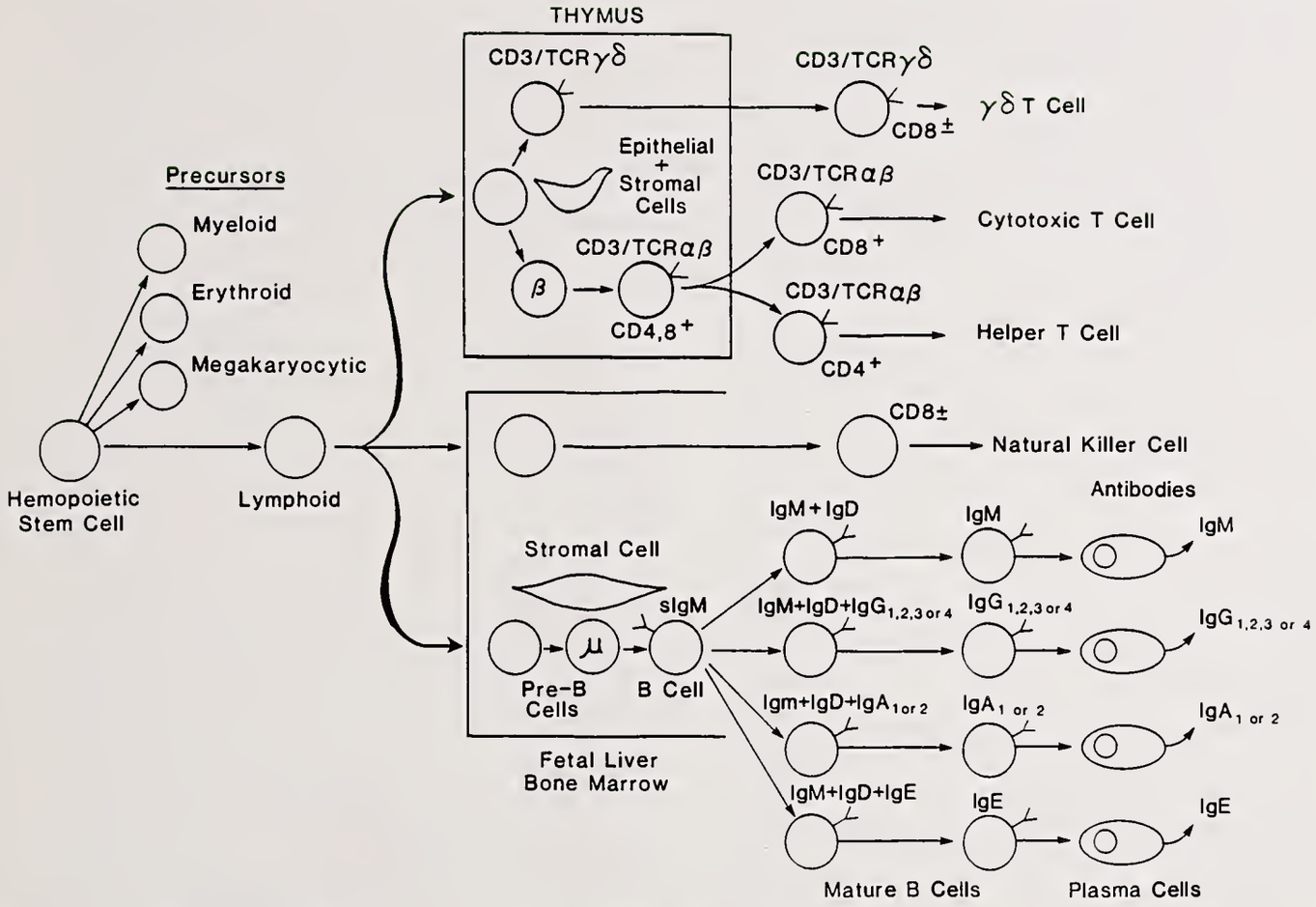
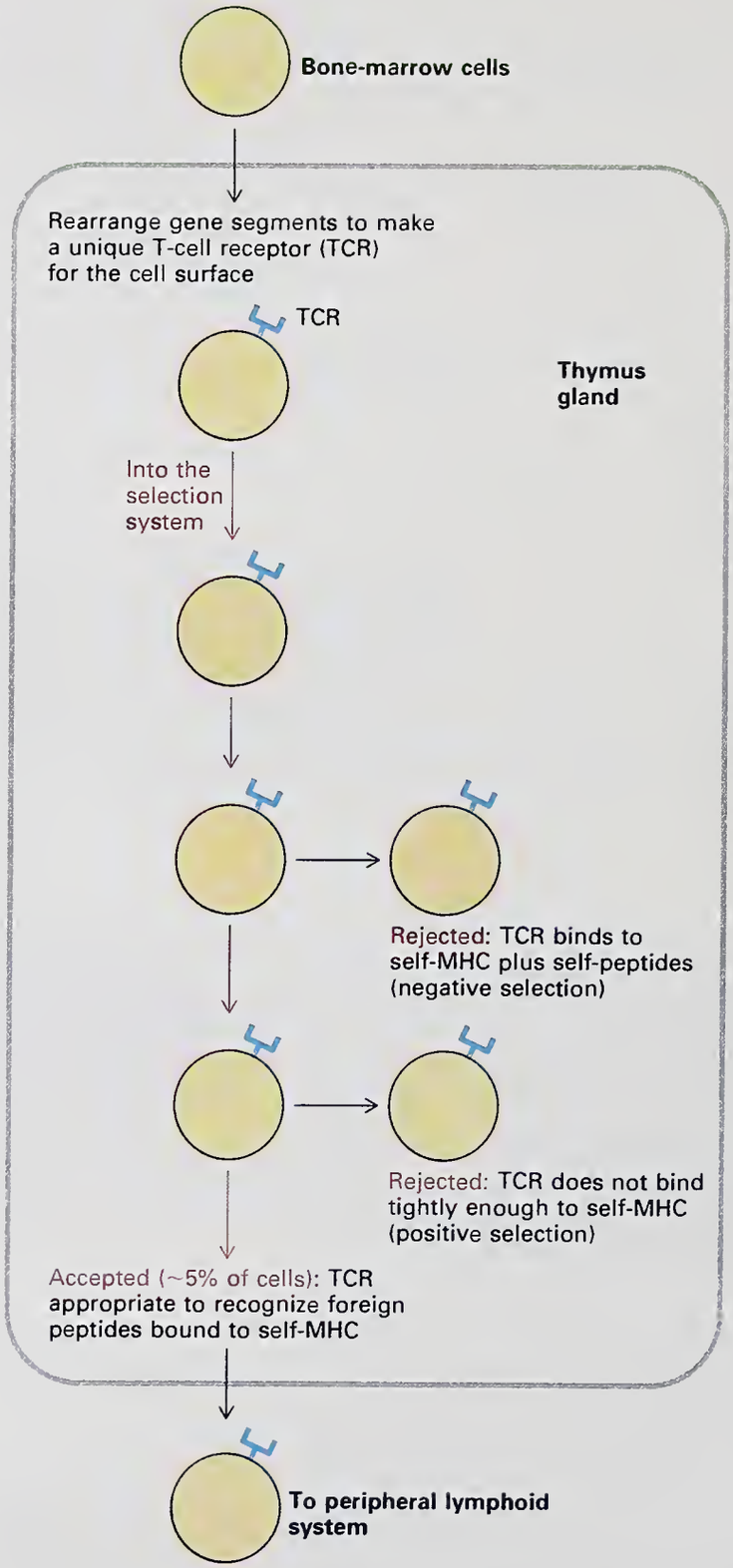


Figure 17. Contemporary model of immune system development in mammals. T and B cells, which form the major recognition and effector limbs of the immune system, are derived from multipotent stem cells that also give rise to the other types of blood cells. Definition of the T cell receptors and accessory 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.*



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 neural transmitter 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 histocompatibility 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 spe-

cialized 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) becomes 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, which ultimately kills the cell. 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 finally the entire T cell population is depleted. Since, as we have seen, these 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 serum and cell membrane proteins (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 more 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 development 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 rela-

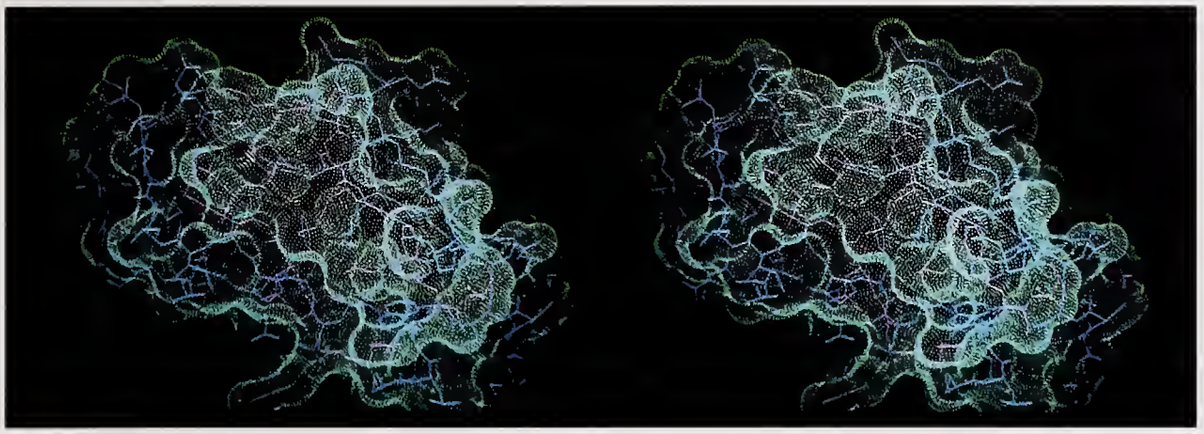


Figure 19. Stereoview of a region on the macromolecule CD4, the T cell receptor for HIV. CD4 contains four tandem immunoglobulin-like domains, D1–D4. The view represents D1's major binding region for the virus coat glycoprotein gp120. Atoms of residues 41–59 are shown in red. They are enveloped by the 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.*

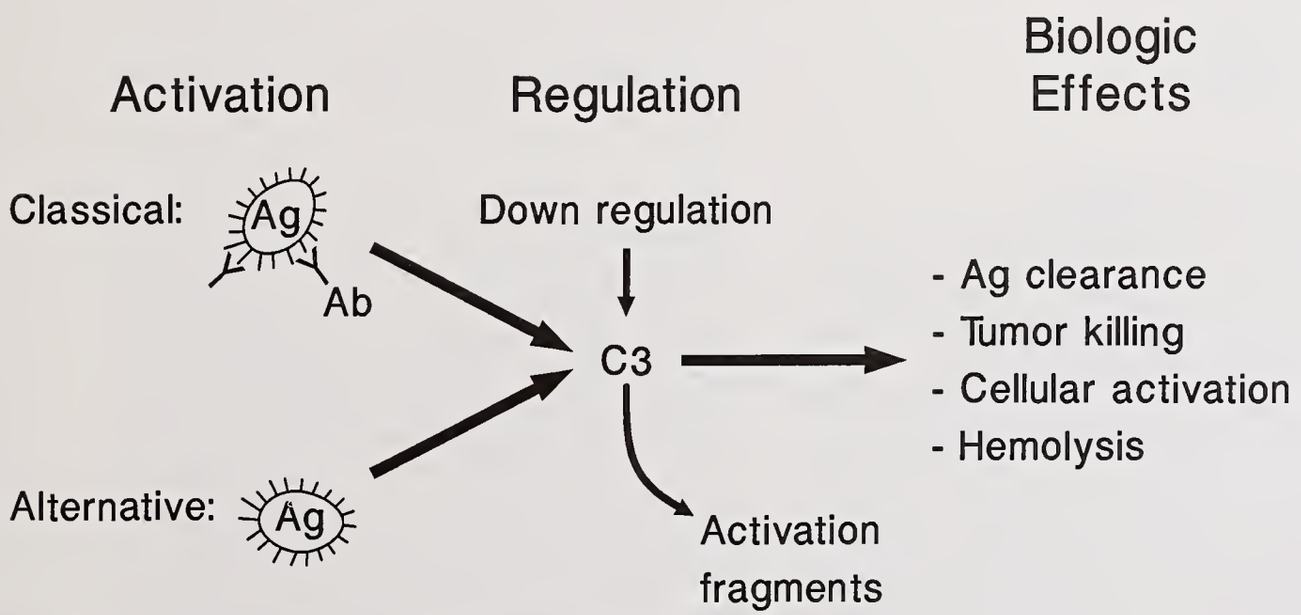


Figure 20. The complement system. Ag, antigen; Ab, antibodies. Courtesy of V. Michael Holers.

tively 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.
Bothwell, Alfred L. M., Ph.D.
Bottomly, H. Kim, Ph.D.
Chaplin, David D., M.D., Ph.D.
Cooper, Max D., M.D.
Davis, C. Geoffrey, Ph.D.
Davis, Mark M., Ph.D.
Fischer Lindahl, Kirsten, Ph.D.
Goodnow, Christopher C., Ph.D.
Greene, Warner C., M.D., Ph.D.
Grosschedl, Rudolf, 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.
Loh, 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.
Rich, Robert R., M.D.
*Schatz, David G., Ph.D.**
Schwartz, Benjamin D., M.D., Ph.D.
Smale, Stephen T., 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.

** This investigator was appointed after manuscripts were submitted for 1991. His research will be described in the next volume.*

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, which, until recently, 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 information 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 300 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 peptides. 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 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

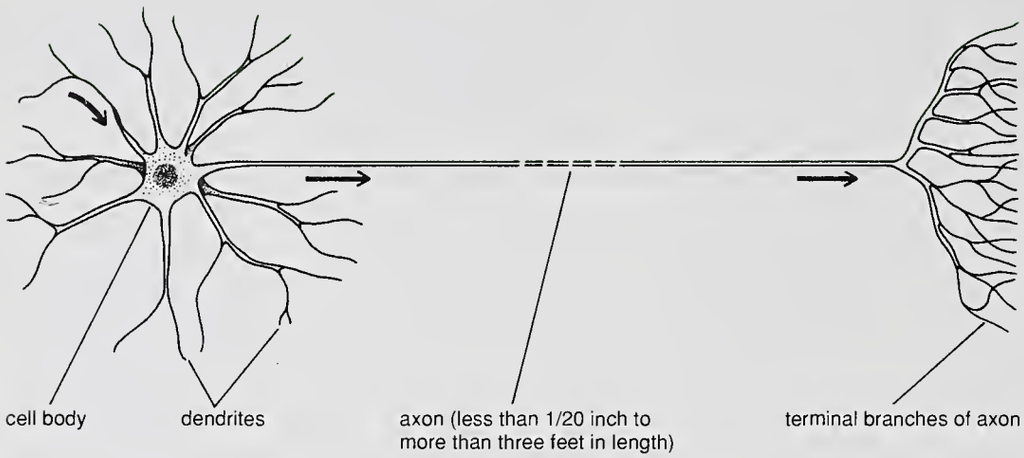


Figure 21. A schematic diagram of a typical neuron. The arrows indicate the direction in which nerve signals are conveyed. The largest axons in the human brain and spinal cord extend for about three feet and have a diameter of less than 1/1,000th of an inch. Many axons are covered by an insulating layer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

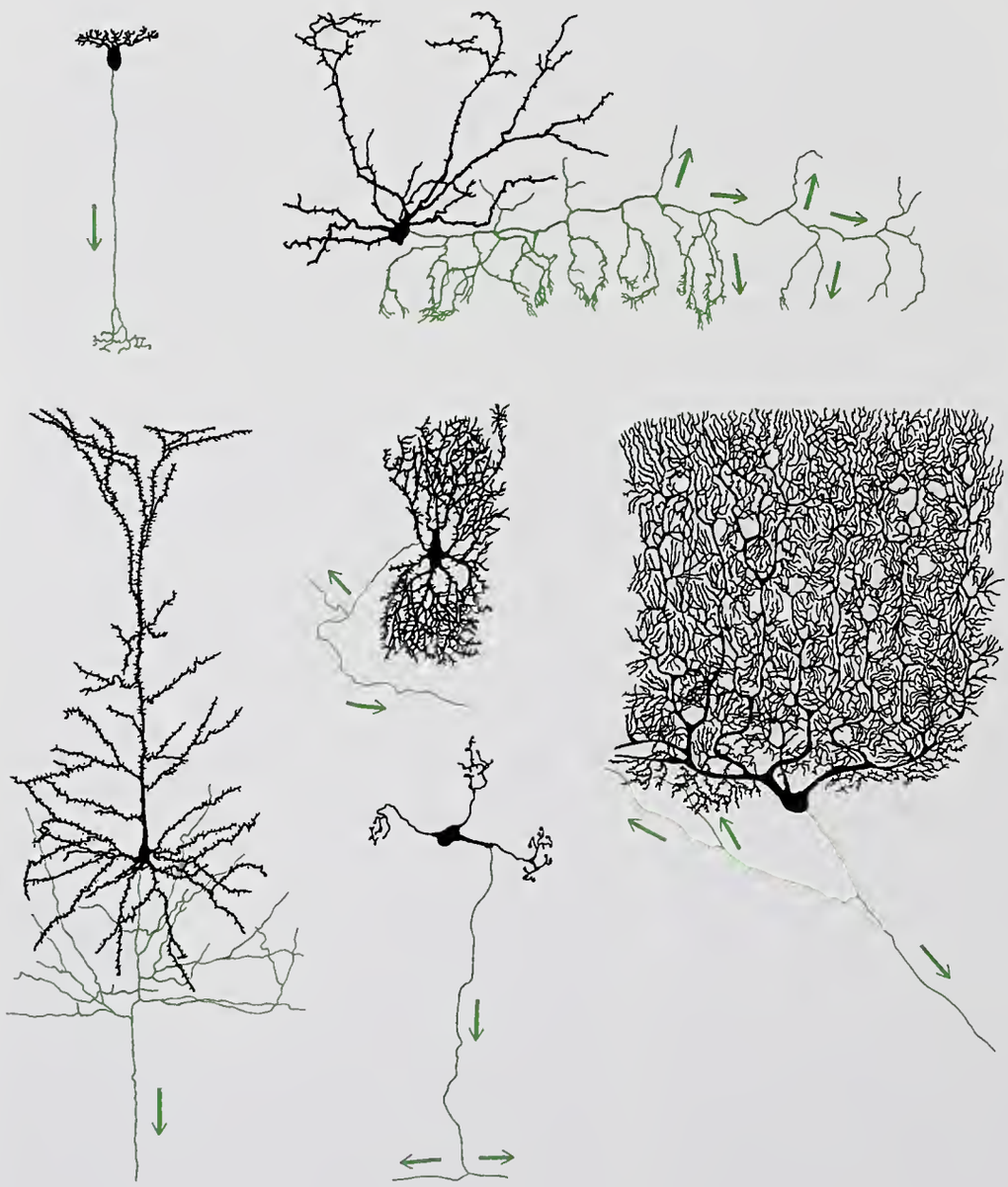
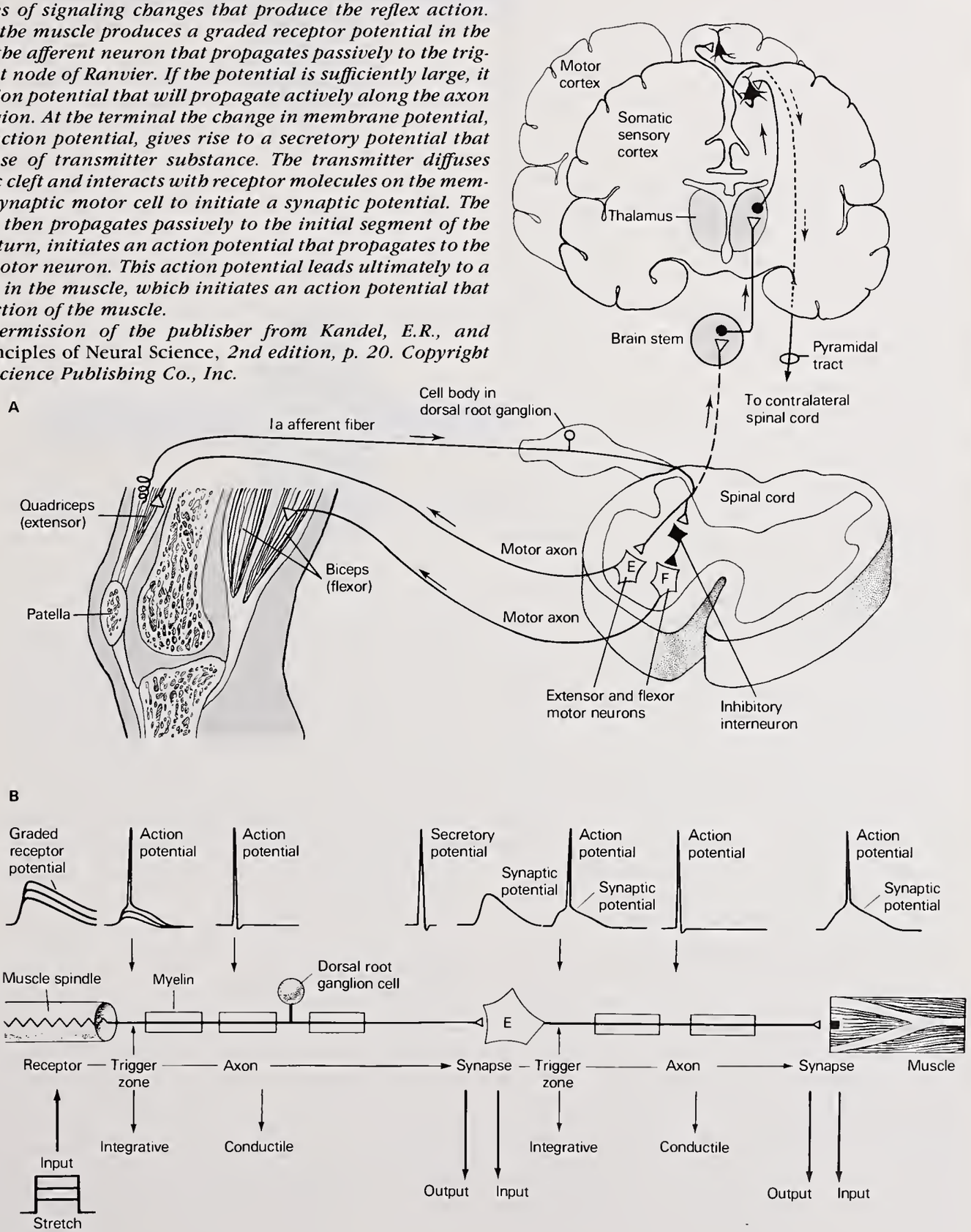


Figure 22. A few of the many types of neurons in the vertebrate nervous system. (S. Ramón y Cajal, Histologie du Système Nerveux de l'Homme et des Vertébrés. *Paris: Maloine, 1901–1911; reprinted, Madrid: C.S.I.C., 1972.)*

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.



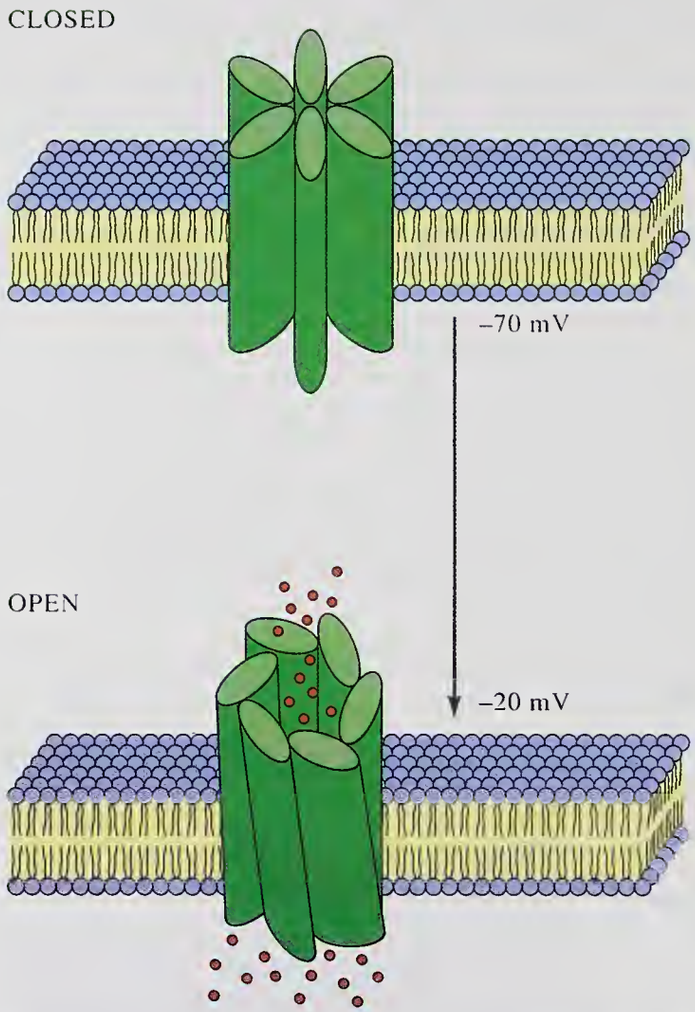


Figure 24. A local reduction of the voltage differential may induce a sodium or potassium channel to change conformation from one allosteric form to another, thus opening the channel and leading to the free passage of ions. This change is presumably caused by key electrically charged amino acids in the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

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.
Amara, Susan G., Ph.D.
Anderson, David J., Ph.D.
Artavanis-Tsakonas, Spyridon, Ph.D.
Axel, Richard, M.D.
Corey, David P., Ph.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.
Jan, Lily Y., Ph.D.
Jan, Yuh Nung, Ph.D.
Jessell, Thomas M., Ph.D.
Kandel, Eric R., M.D.
Katz, Flora N., Ph.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.
Schwartz, James H., M.D., 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.
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.

** These investigators were appointed after manuscripts were submitted for 1991. Their research will be described in the next volume.*

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 func-

tional 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.

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.

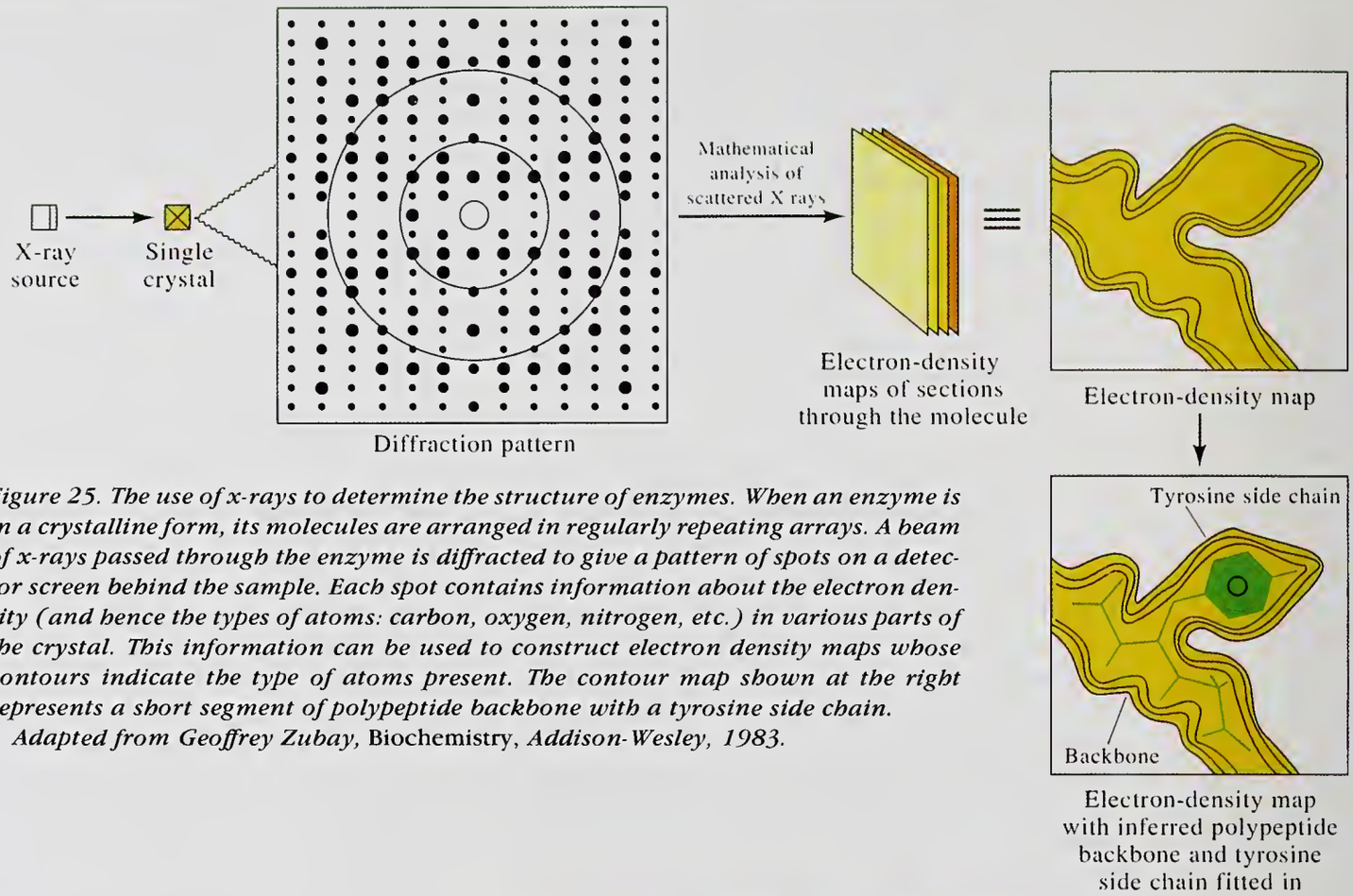


Figure 25. The use of x-rays to determine the structure of enzymes. When an enzyme is in a crystalline form, its molecules are arranged in regularly repeating arrays. A beam of x-rays passed through the enzyme is diffracted to give a pattern of spots on a detector screen behind the sample. Each spot 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.*

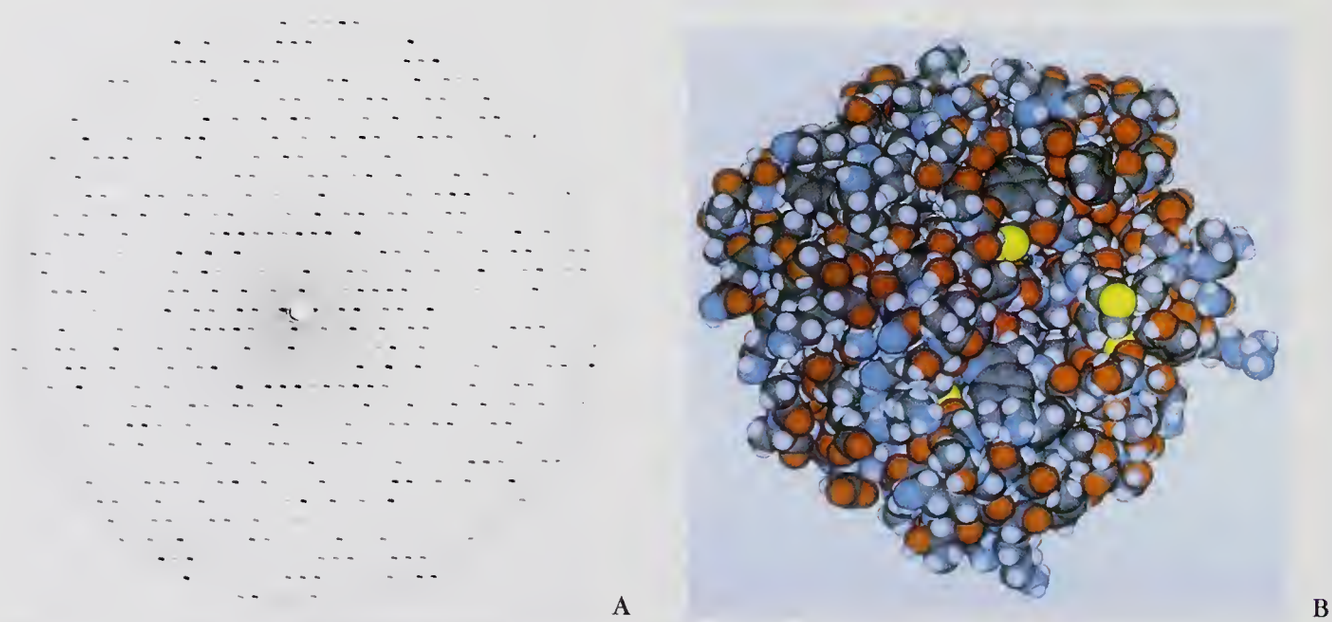


Figure 26. A: The x-ray diffraction pattern given by the enzyme chymotrypsin.

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 Polygen Corporation, Waltham, MA.

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.

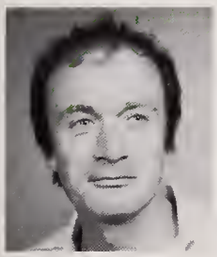
Nuclear magnetic resonance (**NMR**) methods offer an alternative route to determining the three-dimensional structures of peptides and small proteins. The DNA-binding domains of transcriptional activators and repressors are good candidates for this type of analysis. The past two years have seen the 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., Ph.D.
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.
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, Pharmacology, and Neurology 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 recently 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.

The basic electrical pulse of a nerve cell is a positive-negative sequence reflecting the rapid opening of sodium channels followed by their closing and the opening of potassium channels. What triggers the opening and closing of the ion channel gates? It has been known for some time that sodium channel opening is triggered by a positive change in the membrane voltage. We have recently shown that an important trigger for potassium channel opening is a brief increase in cytoplasmic calcium just beneath the cell membrane. Some of our most recent work has focused on how this cytoplasmic signal is generated.

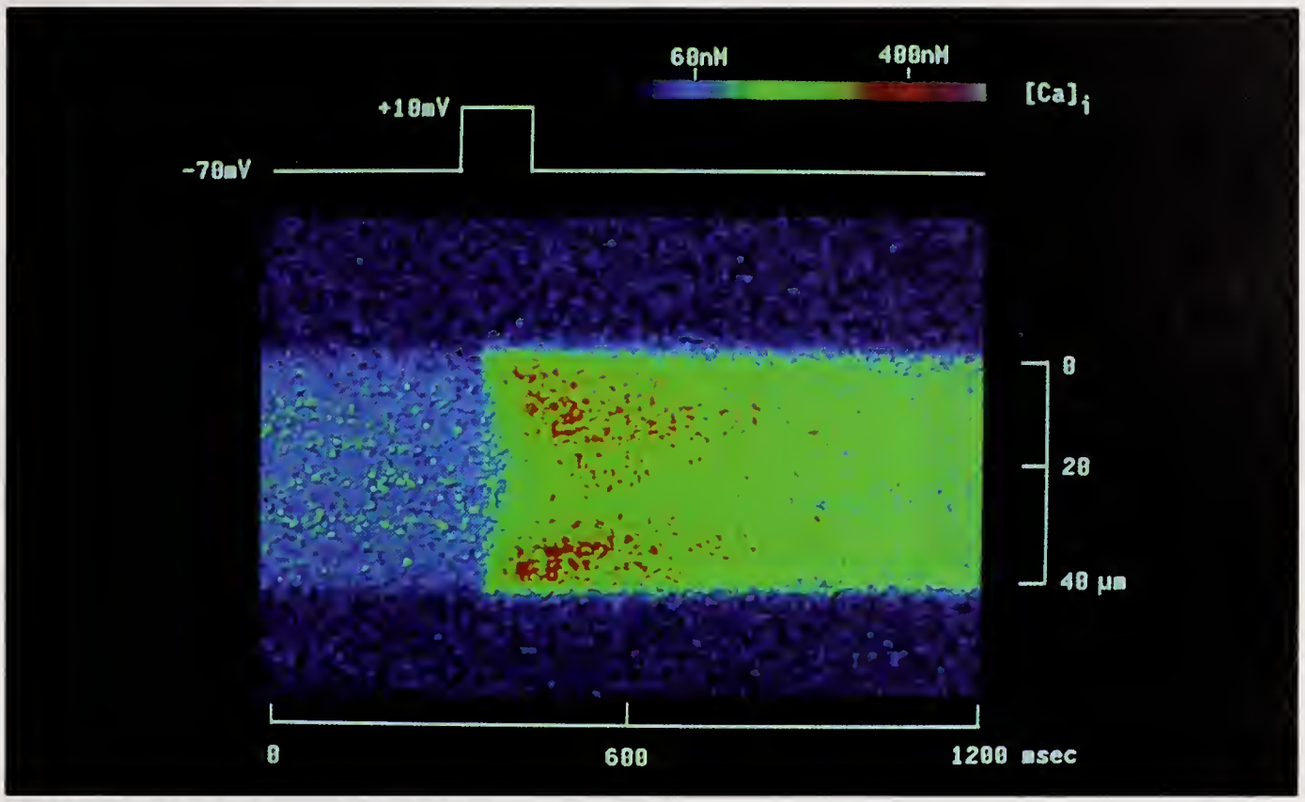
We have split the problem into two parts, using individual nerve cells isolated from bullfrogs. First we analyzed how calcium gets through the membrane from the outside. We employed the voltage-clamp technique, in which an electrical connection is made to the cell interior via a glass micropipette; this allows application of voltages to the cell and measurement of ion movements through channels. The conclusion from this study is that the cell membrane contains numerous channels specialized to allow calcium entry—channels that open quickly when the transmembrane voltage becomes positive and stay open as long as the voltage remains positive. Only one type of calcium channel seems to be operating (other scientists report a more complicated process in other tissues).

The second part of the problem was to determine how calcium spreads within the cell. Here we used a calcium-sensitive fluorescent dye and a scanning laser microscope. The laser beam is focused to a small spot, which can be rapidly scanned over the cell. For example, the spot can be scanned from membrane to membrane in a few thousandths of a second. We find that calcium moves quickly through the cytoplasm once it has entered through the open calcium channels.

Computer calculations show that the movement of calcium can be accounted for by the familiar physical process of diffusion, without any need for special mechanisms. Calcium movements, however, are modified to some extent by binding to cytoplasmic molecules (such as proteins, or the dye we introduced into the cell to measure calcium).

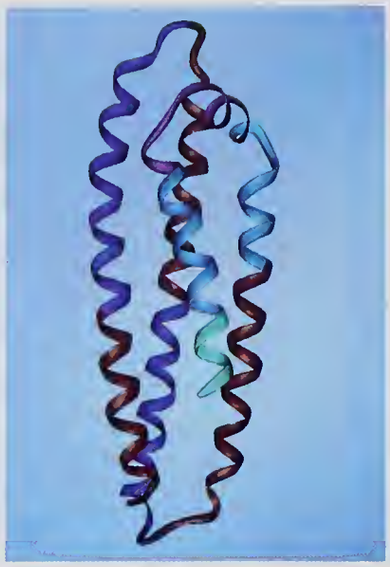
We are also interested in other nerve cell ion channels that are not primarily regulated by calcium. For example, we have recently been able to record directly the activity of a potassium ion channel that we had previously postulated and called the M channel. This channel operates in a membrane potential range (−60 to −30 mV) in which calcium entry is not important. Nevertheless, agents (e.g., the neurotransmitter acetylcholine and the widely consumed drug caffeine) that affect intracellular calcium can also affect the M channel. In recent experiments we have used two different methods to raise intracellular calcium and explore this relationship. One method uses light to release calcium from a molecular cage; the other involves perfusing the cell interior with various calcium-buffer mixtures. Both approaches show that M channel activity increases when calcium levels increase moderately above normal. Transmitters that decrease M channel activity must therefore work by some other, unknown, mechanism.

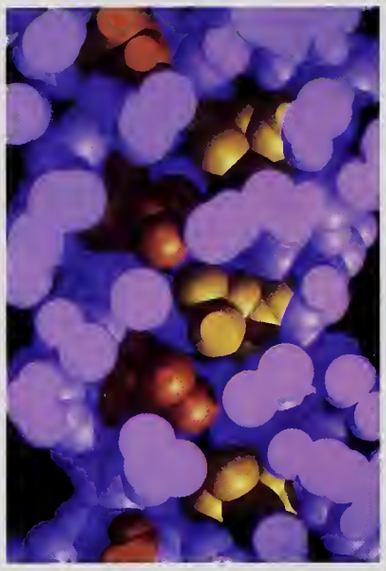
We have also recently studied a neurotransmitter-operated synaptic channel in relay cells of the lateral geniculate nucleus (LGN). Excitatory synapses on LGN relay cells from both the eye and cerebral cortex activate voltage-dependent NMDA (*N*-methyl-D-aspartate) channels. We are currently trying to understand the voltage-dependent responses of these cells.



Calcium diffusion in a nerve cell body. Pseudocolor is used to represent calcium ion concentration at various points along a diameter of a spherical nerve cell. The top edge of the image corresponds to the top edge of the cell (0 µm), and the bottom edge of the image to the bottom edge of the cell (40 µm). Time is represented from left to right. Thus at 0 time, all points along the cell's diameter had a low calcium level (blue). Calcium remained low until the cell was given a voltage pulse (diagrammed at top of figure), which caused calcium channels in the cell membrane to open. Calcium immediately entered the edge of the cell (yellow and then red) and rapidly spread toward the center. The color bar shown is only approximate. Image obtained by combined whole-cell recording and FLUO confocal fluorescence microscopy.

Modified from Hernández-Cruz, A., Sala, F., and Adams, P.R. 1990. Science *247:858–862. Copyright 1990 by the AAAS.*





1 *2* *3*

The receptor-binding domain of apolipoprotein E, the first important protein involved in cholesterol metabolism to be solved by x-ray crystallography. This protein is responsible for targeting plasma lipoproteins such as chylomicrons, VLDL (very low density lipoprotein), and HDL (high-density lipoprotein) to their appropriate cellular receptors (the LDL receptor).

1. Electron density map showing two α-helices from the recently determined crystal structure of apolipoprotein E.

2. Diagram showing that apolipoprotein E is an unusually elongated, antiparallel, four-helix bundle, with an additional short helix linking helices 1 and 2. The structure is built from repeating sequence elements shown by different colored regions.

3. A cutaway surface revealing the hydrophobic interface between helices 1 and 4. The structure is substantially stabilized by a leucine zipper motif. The juxtaposed leucines from one helix are shown in gold, the other in orange.

Research of David A. Agard.

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 Fred 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.

THIS laboratory is primarily concerned with exploring the fundamental relationships between structure and function at the molecular and cellular levels. Four areas of investigation are actively pursued: three-dimensional analysis of diploid chromosome structure and topology; studies on the structural determinants of specificity, using α-lytic protease as a model system; functional and structural analysis of the role of the precursor in proper folding of α-lytic protease; and determination of the first three-dimensional crystal structure of an important protein in human cholesterol metabolism—apolipoprotein E.

Three-Dimensional Analysis of Chromosome Structure

We study chromosome structure in close collaboration with John Sedat (HHMI, University of California, San Francisco); 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 eukaryotic chromosomes as a function of both transcriptional state and the cell cycle stage. To accomplish this goal we are attempting to understand 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 have had to develop the necessary technologies (hardware and software) to allow us to examine complex noncrystalline specimens in three dimensions, using electron microscopy (EM) and light microscopy. The past year has seen significant software developments for three-dimensional image reconstruction with both kinds of microscopes. Many of the tedious aspects of EM tomography have now been automated, greatly speeding the task of generating a three-dimensional reconstruction.

The Role of Topoisomerase II in Chromosome Structures

Current work focuses on deepening our understanding of chromosomal structure by combining three-dimensional observation methods with biochemical probes in an effort to correlate structural aspects with specific macromolecular components. Topoisomerase II is a major chromosomal protein that is important for relieving the torsional stress of supercoiling. It is also postulated to play a crucial role in organizing the higher-order structure of chromosomes. We have begun to investigate the localization and function of topoisomerase II in chromosomal samples prepared by methods that preserve their *in vivo* structure. We have examined prophase, metaphase, and anaphase embryonic cycle 12 and 13 nuclei, and metaphase and anaphase Kc nuclei.

Our data show that although topoisomerase II is clearly associated with mitotic chromosomes, it is concentrated at specific sites along them rather than localized to an internal core. These sites can be spatially coincident with the chromosome or adjacent to the chromosome arm. This seems incompatible with a purely structural role for the enzyme. The nature of the topoisomerase II sites is still unknown, but we hypothesize that those we have recorded are locations of the enzyme's activity during chromosome condensation and segregation. We have observed that topoisomerase II is specifically localized to regions of nondisjunction in failed mitoses.

We are extending this work by examining the distribution of the enzyme at high resolution in the electron microscope and its dynamics in live *Drosophila* embryos. We have recently developed several monoclonal antibodies to topoisomerase II as an important aid to our studies.

The Structural Basis of Enzyme Specificity

We had previously chosen α-lytic protease as a model for investigating structural and energetic aspects of enzyme-substrate specificity, because its binding pocket, made of the side chains of three amino acids (Met 192, Met 213, Val 217A), provided a large volume that could be experimentally manipulated. To probe the structural basis for specificity, we are combining solution kinetic

analysis, x-ray crystallographic structural analysis, and site-directed mutagenesis. Of key importance has been the availability of tight-binding peptide boronic acids, which provide an excellent model for the reaction transition state or nearby intermediates. Approximately 30 high-resolution, extremely well-refined crystal structures have now been determined and analyzed. These have provided significant new insights into the structure of the transition state and the importance of substrate hydrogen bonding for its stabilization, as well as fundamental information on steric exclusion, substrate specificity, and enzyme flexibility.

By mutation we have been able to alter dramatically the pattern of specificity while maintaining or even increasing enzyme activity. Detailed structural analysis of two mutants as free enzymes and as complexes have provided surprising insights into the mechanism and complexity of specificity and have indicated the crucial role of protein flexibility in selectivity. During the past year we have made numerous other mutations and examined their kinetic and structural properties.

In the past year we began a collaboration with Vladimir Basus (University of California, San Francisco) to perform a complete two-dimensional nuclear magnetic resonance (NMR) analysis on the structure of α-lytic protease. We hope this will provide new insights to correlated motions within the enzyme and be useful as well for the folding studies mentioned below.

Recently we have 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 predictions have been remarkably accurate.

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 absolutely required for the proper folding of the 198–amino acid protease domain. Furthermore, we have shown that proper folding can be accomplished either *in vivo* or *in vitro* with the pro region covalently attached or in trans. Our current thinking is that a high-energy barrier exists between the folded and unfolded states that the mature protease cannot cross by itself. The precursor acts as a "foldase" to stabilize the transition state for folding, essentially serving as a template on which the mature enzyme finds its active conformation.

Current efforts are focused on analyzing this folding reaction *in vitro*. Amazingly, we have been able to trap and purify a stable folding intermediate that is rapidly refolded upon addition of the pro region. We plan to use a combination of physical approaches including two-dimensional NMR to probe the structure of the intermediate and the role of the pro region in the final stage of folding.

Structure of Apolipoprotein E

Apolipoprotein E is an important protein in cholesterol metabolism in mammals. It is one of two proteins that can bind to the low-density lipoprotein (LDL) receptor (the other is apolipoprotein B) and thus 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, whereas 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 have obtained crystals of the 22-kDa receptor-binding domain and recently finished the high-resolution structure determination. The protein is an unusually elongated four-helix bundle. Although the surface is exceptionally charged, positive and negative charged groups are precisely balanced, except in what we believe is the receptor-binding region. Currently, we are examining the structures of two human mutants that disrupt receptor binding.

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 and the Young Investigator Award of the Society for Neuroscience.

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 the integration of these signals to generate appropriate behavior. In addition, ion channels are important for the generation and regulation of the heart beat, for contraction of muscles, and for the release of hormones in the bloodstream. A 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 respond 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 their transmission throughout the body. These molecules have a way to measure the electrical potential and open accordingly. In addition, some of them 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 recombinate DNA methods and expressing the normal and altered channels in frog oocytes. Our results demonstrated that the first 10 or so hydrophobic amino acids, and the positively charged amino acids in positions 16 through 19, are important in the inactivation mechanism. By altering these amino acids, we can change the inactivation rate in a graded manner up to about 20 times slower than normal.

Deletion and insertion mutations in an adjacent region alter the inactivation rate as a function of the length of the remaining amino acid chain, suggesting that this region acts as a spacer. 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 agonist for inactivation. The inactivation receptor is located elsewhere on the cytoplasmic side of the molecule. When the inactivation agonist is bound to the receptor, the channel closes. The first 10 or so hydrophobic amino acids of the *Shaker* protein either form the core of the inactivation particle or provide the hydrophobic interactions necessary for the particle to bind to the inactivation receptor. The positively charged residues electrostatically interact with the receptor, which is likely to be negatively charged. Mutations in this region presumably disrupted the inactivation particle and precluded inactivation.

Amino acid residues 21 and beyond form the

"chain." Deletion mutations in this region accelerated the inactivation process by shortening the chain part of the sequence. We tested this model further by applying a solution containing the 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. We are now using synthetic inactivation particles and mutant channels to examine the detailed structural requirements for inactivation.

Shaker potassium channels also exhibit a slower inactivation process. It 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 (HHMI, the Johns Hopkins University) we have found that the slow inactivation is affected by external agents, suggesting that the conformational changes for this process involve external structures.

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 to a single hydrophobic amino acid in a membrane-spanning region of the channel molecule the difference responsible for the slow inactivation differences. Other amino acid substitutions at this position have dramatic effects on gating, with larger hydrophobic amino acids leading to slower inactivation. We are currently investigating this process further.

Molecular Mechanisms of Tissue-Specific Hormonal Regulation of Gene Expression

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

Dr. Alexander-Bridges is also Assistant Professor of Medicine at Harvard Medical School and Clinical Assistant at Massachusetts General Hospital. 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.

ONGOING studies in our laboratory are aimed at elucidating the mechanism of insulin action on the expression of enzymes that regulate cell growth and metabolism. We have focused particularly on regulation of the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene in adipose tissue and liver. Insulin induces GAPDH mRNA levels 8-fold in cultured 3T3-L1 adipocytes and 10-fold in fat or liver tissue isolated from rats fasted and then fed a high-carbohydrate, low-fat diet. Expression of the GAPDH gene is markedly decreased in primary adipocytes isolated from diabetic animals and induced above basal levels upon replacement of insulin. The effect of insulin on this gene is tissue specific; GAPDH mRNA is not regulated in muscle. Study of insulin regulation of this gene provides a marker of the metabolic effects of insulin on gene expression.

Activation of GAPDH gene expression in insulin-responsive tissues correlates with the presence of insulin-responsive DNA-binding proteins that bind specifically to elements in the 5′ flanking region of the GAPDH gene and confer insulin-responsive gene expression to the chloramphenicol acetyltransferase gene. Within 60 minutes of exposure of 3T3 adipocytes or H35 hepatoma cells to insulin, the activity of these sequence-specific DNA-binding proteins is increased two- to fourfold. The insulin-responsive element A (IRE-A) DNA-binding protein (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, a process known to increase circulating glucose and insulin levels, resulting in the induction of glycolytic and lipogenic enzymes. IRP-A binding is inhibited in the fat pads of diabetic animals and is induced above normal levels when insulin is administered to diabetic animals. In muscle, where GAPDH activity is not rate limiting, IRP-A binding is not detectable. These observations support the importance of GAPDH gene regulation *in vivo*.

The Southwestern screening technique was used to clone a gene that encodes a specific IRP-A DNA-binding protein. The cloned cDNA has been used to examine the mechanism by which insulin chronically regulates expression of the GAPDH gene. This clone is expressed in liver and fat, but not in muscle, which provides an explanation of the tissue-specific regulation of GAPDH gene expression. Expression of IRP-A mRNA is inhibited in diabetes and up-regulated with insulin replacement; expression is induced during the process of fasting and refeeding. In contrast, 1 hour of insulin exposure of cells does not appear to alter expression of the IRP-A mRNA. Thus it appears that the activity of this factor is regulated acutely by a post-translational modification and chronically by an alteration in gene expression.

The mechanism by which insulin's initial interaction with its cell surface receptor tyrosine kinase stimulates intracellular processes is unclear. Several enzymes that are regulated by insulin have been found to undergo phospho-dephospho interconversions. Binding of IRP-A protein to DNA is dependent on phosphorylation. To define the steps in the signal transduction pathway of insulin action on this transcription factor, we will use antibodies that specifically interact with IRP-A protein and examine the effect of insulin on the phosphorylation state of the IRP-A protein. IRP-A contains a binding domain that is perfectly conserved across species from rat to yeast and an acidic domain that is capable of activating gene transcription. These domains are surrounded by putative phosphorylation sites for insulin-sensitive kinases. Future efforts will be aimed at determining whether any of these putative phosphorylation domains is critical for activation of DNA binding or transcriptional activity.

Southern analysis with the conserved IRP-A binding domain indicates that this gene belongs to a large family of related genes. Over the next year, the potential for interaction between family members will be explored. In the process, the regions of IRP-A that contact DNA and the regions that are required for protein-protein interactions will be defined.

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 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 Biochemistry and Microbiology at Columbia University College of Physicians and Surgeons. 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. 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). B and T lymphocytes that actively fight infections and other diseases are generated in two general stages (primary and secondary).

During primary stages, 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 secondary phase of lymphocyte differentiation results when a lymphocyte meets a foreign antigen (e.g., a bacterial cell surface component) that is recognized by its surface receptor; this stimulates the lymphocyte to divide and mature into an effector cell. For B lymphocytes, this maturation event involves secretion into the bloodstream of its specific receptor molecule; this secreted product is the antibody.

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. 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 their basic cassettes, 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.

One experimental system employed in our studies involves the generation of cell lines that undergo the same types of antibody gene assembly events in culture dishes that normal developing immature lymphocytes undergo in the animal. In particular, we have devised and employed assays in which unrearranged antibody gene segments are isolated from normal mouse or human chromosomes and then propagated (cloned) in bacteria; we refer to these cloned molecules as "recombination constructs." Various types of recombination substrates can be introduced into the permanent cell lines to ask mechanistic and regulatory questions about antibody gene assembly. These studies have allowed us to clarify many mechanistic aspects of this gene assembly process and to devise schemes for how the assembly process is regulated. For example, we were able to show that both antibody genes and the analogous genes in T lymphocytes (encoding the T cell receptor) are assembled by the same general recombination machinery.

A second general system that we employ to analyze control of the recombination mechanism is the transgenic mouse. By injecting purified genetic material into fertilized mouse eggs and then implanting them into foster mothers, it is possible to make mice that carry the foreign DNA in their germline. We have inserted recombination substrates that contain experimentally manipulated antibody gene cassettes (or T cell receptor gene cassettes) into the mouse genome and asked questions about the genetic and cellular mechanisms that regulate assembly. These studies allowed us to extend our studies on clonal cell lines to identify more precisely several genetic

elements that govern the assembly process. Most notably, we have shown that genetic regulatory elements previously thought to control only the level of gene expression can also serve to control specific gene rearrangement.

Currently we are attempting to generate several novel mouse models to study factors involved in lymphocyte differentiation. One approach is to produce a mutant mouse line that cannot produce endogenous antibody molecules. To accomplish this, we employed new methods that permit introduction of desired genetic mutations into specific genes in mouse embryonal stem cell lines to create cells that lack germline gene segments necessary for forming functional antibody genes. We used these mutant cell lines to create chimeric mice that contain cells lacking the antibody gene cassettes. Such chimeric mice should generate mutant mice that inherit in their germline an inability to make endogenous antibodies.

We will use the mutant 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 a transgenic mouse line that we have engineered to contain functional human antibody gene cassettes. The hope is that the hybrid animal will now rely on the human antibody genes for its immune system, providing a more effective method of generating tailor-made human antibodies.

We also are interested in the mechanism by which certain intracellular messenger molecules (lymphokines or interleukins) affect the development and action of B lymphocytes. Certain T lymphocytes regulate aspects of the immune response; some of these produce lymphokines that stimulate activated B lymphocytes (during secondary differentiation) to produce antigen-specific antibodies that differ in where they go in the body and how they mediate their protective activity (referred to as different effector activities). The difference in these antibodies results from the use of different gene cassettes to encode, not the antigen-binding portion, but the effector portion of the antibody molecule. This phenomenon results from a different type of gene recombination assembly event, in which one part of an already active antibody gene is replaced with a different one. We have shown that particular lymphokines actually direct this gene rearrangement event by giving a B lymphocyte a signal to prepare a specific effector gene cassette for a rearrangement event. Most recently we have shown that the ultimate targets of the lymphokines are novel elements that control the transcription of the effector gene cassette.

Finally, we are interested in the molecular signals that direct precursor lymphocytes through the primary differentiation process. One focus of this work is on three related but distinct genes, the N-, L-, and c-*myc* genes. These genes encode proteins that operate in the nucleus of the cell, presumably to regulate the expression of other genes. The members of this gene family contribute to the generation of various types of tumors when their expression is altered as the result of mutations. We have found that individual members of this family are expressed at specific points during the differentiation of lymphocytes. For example, the N-*myc* gene is expressed only in lymphocyte precursors where expression levels are greatly increased following interaction of these cells with a factor that specifically stimulates their growth.

To define further the function of this gene family, we have introduced the N-*myc* and L-*myc* genes into transgenic mice in forms in which they are incorrectly expressed throughout lymphocyte development (normally they are expressed only very early in lymphocyte development). Such incorrect expression of these two genes leads, respectively, to defects in B or T lymphocyte development, including frequent generation of lymphocyte tumors. The genetic mechanisms by which incorrect expression of the deregulated *myc* genes affects normal lymphocyte development are being analyzed.

As a complementary approach, we used the gene-targeting methodologies outlined above to create mutant mice that lack functional N-*myc* genes. Loss of N-*myc* gene function appears to be lethal at an early stage of murine development. Therefore, to study normal function, we are attempting to rescue development in the N-*myc* mutant mice lacking functional endogenous genes by breeding these animals with mice that carry functional N-*myc* transgenes that are expressed in various cell lineages.

Molecular Studies of Neurotransmitter Transport and Regulation of Neural Gene Expression

Susan G. Amara, Ph.D.—Assistant Investigator

Dr. Amara is also Assistant Professor of Pharmacology at Yale University School of Medicine. She received her B.S. degree in biological sciences from Stanford University and her Ph.D. degree in physiology and pharmacology from the University of California, San Diego, where she worked with Michael Rosenfeld. Dr. Amara began her postdoctoral studies at San Diego, in the Eukaryotic Regulatory Biology Program. She continued these studies at Yale as a Fellow of the Life Sciences Research Foundation.

COMMUNICATION between neurons depends on precise chemical signals that are released from one neuron and interpreted by adjacent neurons. This transfer of information is based on the synthesis of chemical neurotransmitters by the presynaptic cell, their release into the synaptic cleft, and their recognition by specific receptor proteins on the membrane of the postsynaptic cell. Our research focuses on the structure and expression of genes encoding proteins with important roles in neurotransmission and synaptic function.

We have been interested in a family of membrane proteins that play a key role in synaptic function—the sodium-dependent neurotransmitter transporters. For precise and efficient neurotransmission, the release of the neurotransmitters into the synaptic cleft must parallel the rise and fall of presynaptic excitation. Inactivation of the majority of classic neurotransmitters usually is achieved by rapid reuptake into the presynaptic terminal or surrounding glial cells by specific, high-affinity neurotransmitter transporters. Neurotransmitter entry is coupled directly to the transmembrane sodium gradient, which drives the uphill transport. Neurotransmitter transporters not only have a central role in synaptic transmission, but they are also the site of action for a wide range of drugs with both therapeutic and abuse potential. Clinically important drugs that act directly on these transporters include the tricyclic antidepressants, which inhibit both norepinephrine and serotonin uptake; amphetamines, which inhibit norepinephrine and dopamine transport; and cocaine, which inhibits all three uptake systems. The impact of these agents on society underscores the importance of understanding the molecular basis of their actions.

Until recently little structural information has been available about the proteins responsible for reuptake of these neurotransmitters at synapses. Previously we used *Xenopus* oocytes as a system in which to express four major classes of transport activity in the brain—those for catecholamine, indoleamine, choline, and excitatory and inhibitory amino acid (L-glutamate, γ-aminobutyric acid [GABA], and glycine) transport. After injection of mRNA prepared from various brain regions, uptake of the radiolabeled transmitters can be measured and the ion-dependence and pharmacologic specificity of these transport activities can be assessed in single oocytes. These and other studies, which confirmed that the transport activities are encoded by single mRNAs, paved the way for expression cloning of the genes encoding these carriers.

This year we utilized a mammalian cell-based expression system to clone a human cDNA encoding a catecholamine transporter. The transport activity encoded by this cDNA has the substrate-specificity and pharmacologic properties of a norepinephrine transporter, including sensitivity to antidepressants, amphetamine, and cocaine. The single polypeptide contains regions important for binding and transport inhibition by cocaine and thus provides a first structural insight into the family of cocaine receptors. The cDNA sequence predicts a protein of 617 amino acids, with multiple hydrophobic regions consistent with the presence of 12 membrane-spanning domains. The amino acid sequence has significant homology with a recently cloned GABA transporter and thus identifies a new family of brain transport proteins. Studies are under way to define and characterize other members of this carrier family.

The availability of a human norepinephrine transporter cDNA also provides an opportunity to investigate the contributions of structural elements to transporter function and regulation. Many tricyclic antidepressants accomplish their therapeutic effects by blockade of reuptake of norepinephrine transporters and subsequent elevation of synaptic neurotransmitter concentrations. Delineation of the structural basis of antidepressant binding and function through mutagenesis studies of the cloned transporter should aid in the development of more-selective therapeutic agents for the treatment of human depression. The availability of the human cDNA encoding the norepinephrine carrier also offers the first opportunity to determine whether alterations in

transporter genes underlie major psychiatric affective disorders.

Other ongoing research in the laboratory has focused on genes encoding another class of important synaptic molecules, neuropeptides. Using the calcitonin/CGRP (calcitonin gene-regulated peptide) gene family as a model system, we are investigating the mechanisms that regulate the production and structural diversity of these peptide neurotransmitters. The peptides of the calcitonin/CGRP family are produced in discrete sets of neurons in the nervous system. We are looking for the specific signal sequences within the genes for these peptides that allow the gene to be turned on in these sets of neurons. In addition we are determining which sequences are involved in regulating the amount of neuropeptide mRNA produced. Once these specific control sequences are identified, our next step is to search for the protein factors that recognize the sequences and influence gene expression. Thus we can use a molecular biologic approach to characterize the individual steps that combine to produce complex neurobiologic processes.

The diversity of products expressed from the calcitonin/CGRP gene family depends not only on the presence of multiple, independently regulated genes but also on the existence of alternative, tissue-specific pathways for pre-mRNA processing. This processing is an important mechanism in the control of the expression of RNA products from many eukaryotic genes. Numerous gene products in the nervous system are encoded within complex transcription units that generate multiple mRNAs by altering where in the gene the transcripts initiate, how they are spliced, or how they terminate. In the nervous system, these processes may be particularly important, because they increase the diversity of expression by allowing multiple products to be encoded by a single gene. We would like to know how the same nuclear RNA transcripts can be processed in alternative ways in different cells to generate distinct products. One potential determinant of a processing choice is found within the primary RNA transcript, either as a specific sequence or secondary structural feature. A second determinant is found in the enzymatic machinery that recognizes these structural features and targets the transcript to a particular processing pathway. We have continued to generate a variety of mutated and hybrid genes and, by introducing these constructs into a variety of cell types, have identified regions critical for splicing regulation.

Control of Cell Fate During Vertebrate Neuronal Development

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

Dr. Anderson is also Assistant 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 for work with Günter Blobel. He then did postdoctoral research in molecular neurobiology with Richard Axel at Columbia University. He counts among his honors a Searle Scholars Award and an NSF Presidential Young Investigator Award.

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 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 culture environment of the cells, 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, 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 indicates 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 some 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.

Immortalization of the Chromaffin-Neuron Precursor Cell

The precursor cell is endowed with a limited repertoire of potential developmental fates. What genes and proteins determine the specific repertoire of possible fates? What genes and proteins select a particular fate for actual expression? These are the major questions we are pursuing in an effort to understand the molecular biology of this developmental system.

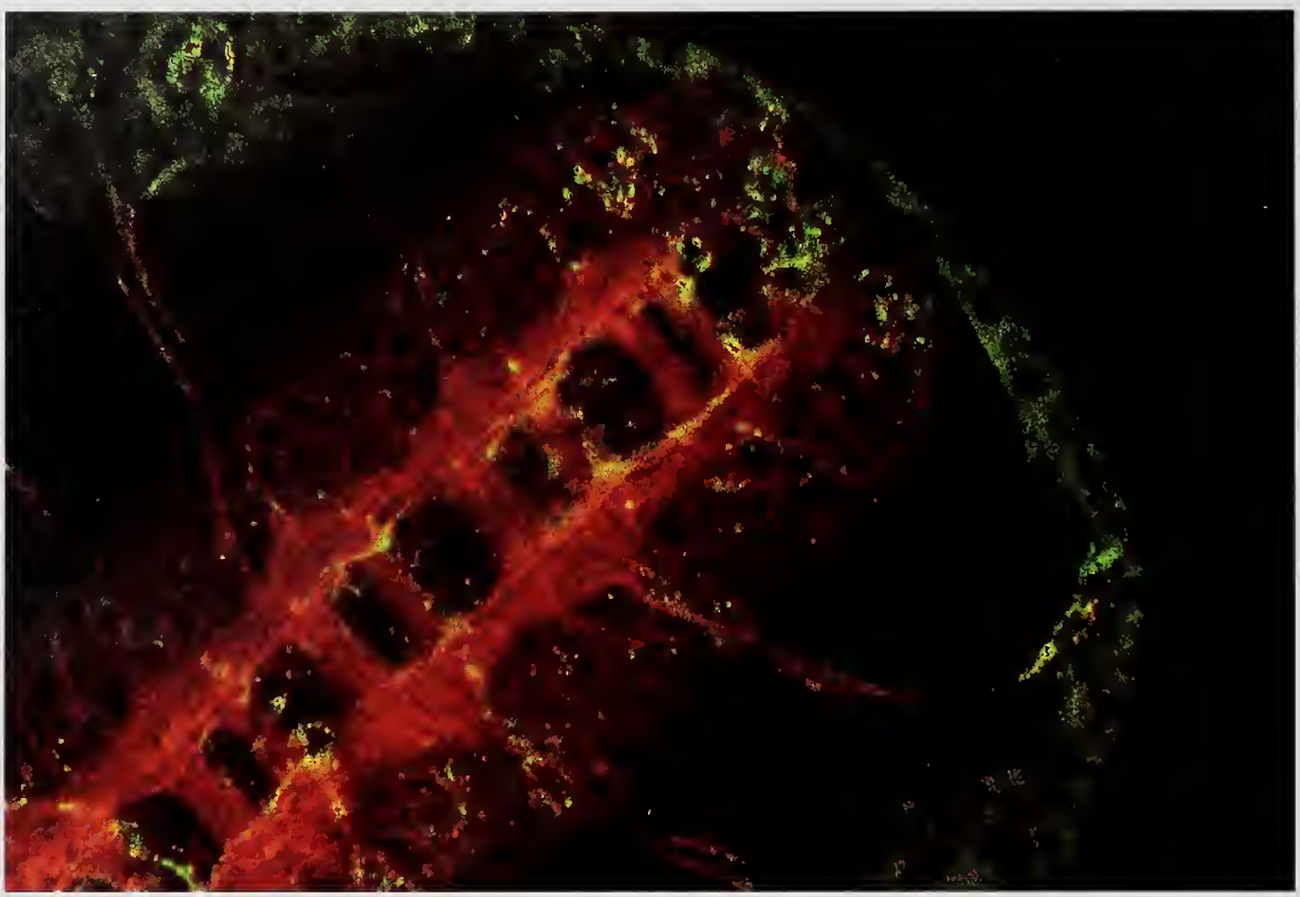
One problem in studying the chromaffin-neuron precursor cell at the molecular level is the small number of cells that can be isolated from rat fetuses. To circumvent this problem, we have applied recently developed techniques to immortalize the cells. Using a defective retrovirus as a "disposable molecular 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.

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. This approach involves first guessing that a particular gene, or class of genes, may be expressed in the precursor cell and then using the tools of molecular biology to extract the gene of interest. What guidelines can one use to make educated guesses about developmental control genes? One suggestion came from studies, by

other investigators, which showed that embryonic control genes identified by mutational analysis in the fruit fly *Drosophila* have counterparts in the mouse and human that are remarkably conserved in their structure. We therefore reasoned by analogy 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. If the function of these genes is eliminated by mutation in the fly, certain classes of neurons fail to form. 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. This discovery opens up a new line of investigation, which will be aimed at finding both the precise role of the *scute*-related genes during neuronal development and the manner in which these genes are controlled. These exciting findings suggest that the molecular mechanisms controlling nerve cell development in vertebrate and invertebrate organisms may be fundamentally similar. Studies in each system will therefore contribute to our understanding of the other, leading to a unified view of the basic principles that guide the building of a nervous system.



Confocal microscopic localization of the JR-8D2 epitope to specific fascicles within the major nerve tracts of Drosophila melanogaster. *Neuronal membranes are shown in red with the staining from the JR-8D2 antibody shown in green. The protein recognized by JR-8D2 exhibits an exquisite localization to both neurons, which pioneer the peripheral nervous system, and to specific fascicles in the central nervous system.*

From Rothberg, J.M. 1991. Ph.D. thesis, Yale University.

Cell Interactions in Neurogenesis

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 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. He joined the Yale faculty in 1981.

A fundamental issue in the development of multicellular organisms is how an individual cell acquires a specific developmental fate. The molecular rules directing neighboring cells into developmentally discrete paths are unknown. With a particular interest in how these rules apply to the nervous system, we have been studying the molecular biology of early neurogenesis in the fruit fly *Drosophila*.

The central nervous system in *Drosophila* derives from a set of precursor cells, the neuroblasts, which segregate from the epidermal precursors, the dermoblasts, in the very early ectoderm. For the past several years, we have been analyzing a group of six zygotically acting genes that are involved in this process. Known as neurogenic loci, these are *Delta* (*Dl*), *Enhancer of split* [*E(spl)*], *mastermind* (*mam*), *big brain* (*bib*), *neuralized* (*neu*), and *Notch* (*N*). Mutations in any one of these can result in the misrouting of epidermal lineage into a neuronal developmental pathway.

The first neurogenic locus to be characterized in some detail was *Notch*. We found that it codes for a transmembrane protein homologous to the mammalian epidermal growth factor (EGF), implying an involvement in cell surface events. Indeed, our work and that of others—embryological and molecular studies of some of the other neurogenic loci—strengthened that notion and led to the hypothesis that neuroblast segregation depends on cell interactions, with at least some of the neurogenic loci coding for proteins involved in intercellular signaling.

In studying the cellular mechanisms in which *Notch* is integrated, we have been following two major experimental paths. On the one hand, we are exploring the cell biology of *Notch* and the nature of its genetic interactions; on the other, we are trying to gain insights into the functional meaning of certain sequence motifs found in the *Notch* protein. For instance, we are characterizing other genes sharing the EGF sequence motif with *Notch*. In addition, during the past year, we initiated a molecular and genetic analysis of certain nonneurogenic genes that play a role in the differentiation of the ectoderm.

We started this analysis with the gene *fizzy* (*fzy*). Mutations of this gene affecting the phenotype have suggested that it functions to promote subsequent differentiation in response to the postulated "epidermalizing" component of the signal transmitted by the neurogenic gene products. As such, the *fzy* protein has a pivotal role in implementing the response of presumptive dermoblasts to developmental cues transmitted by those products. We have been analyzing the genetic and phenotypic behavior of *fzy* mutations and have initiated its molecular analysis by the cloning of genomic sequences from the *fzy* region.

In an attempt to examine the complexity of the genetic circuitry in which *Notch* is integrated, and to identify genes whose products may directly interact with the *Notch* protein, we designed a genetic screen aimed at identifying suppressors of certain *Notch* mutations associated with the gene's extracellular EGF homologous domain. This screen has led to the identification of a surprisingly restricted set of interacting loci. These include the neurogenic genes *Delta*, which also codes for a transmembrane protein with EGF homology, and *mastermind*. In addition, a third gene was shown capable of acting as a suppressor. We identified this gene to be the *deltex* locus, hitherto unlinked with neurogenic gene action. We have further characterized the phenotype of *deltex* mutations and have demonstrated allele-specific interactions between *deltex* and *Notch* alleles.

To investigate the possibility of intermolecular association between the products of *Notch* and *Delta* genes, we have examined the effects of their expression on aggregation in *Drosophila* S2 cells. We find that *Notch*-expressing cells, by a calcium-dependent process, form mixed aggregates specifically with cells that express *Delta*. Furthermore, we have determined that *Notch* and *Delta* proteins interact via their extracellular domains.

Some of the first phenotypic interactions to be described between neurogenic loci were those involving *Notch* and *Enhancer of split*. Molecu-

lar and genetic analysis of *E(spl)* revealed that the locus is defined by several transcription units. Focusing mostly on one of those transcripts, we have shown that it codes for a protein with striking homology to the bovine β-transducin, known to be involved in signal transduction. We found that point mutations in the protein are capable of dramatic synergistic interactions with mutations affecting the intracellular region of *Notch*. We were rather surprised to find the subcellular localization of the *E(spl)* transducin homologous protein in the nucleus. *mastermind*, which also interacts with *Notch*, seems to code for a nuclear protein as well. This raises the intriguing possibility that *Notch* communicates with the nucleus relatively directly.

While examining identified interactions between *Notch* and other genes, we are studying two more, *Serrate* and *slit*. The *Serrate* locus captured our attention by virtue of the synergistic genetic interactions we noted between certain alleles of *Serrate* and *Notch*. We were particularly intrigued to find, after cloning *Serrate*, that it encodes a protein product of 1,404 amino acids with a single transmembrane domain and 14 EGF-like repeats in the extracellular region. Thus *Serrate* represents another member of the group of EGF-containing loci in *Drosophila* and provides us with the reasonable working hypothesis that *Notch* and *Serrate* may interact at the protein level.

To gain more insight into the role of the EGF sequence motif in extracellular interactions and morphogenetic events, we have undertaken a molecular and biochemical characterization of *slit*. This gene codes for a secreted protein with sequences homologous to EGF and to a superfamily of extracellular matrix–binding glycoproteins. Its embryonic localization, mutant phenotype, and sequence homology suggest that it mediates interaction among glial cells, axons, and the extracellular environment. The interactions include those required for developing cells that are necessary for axon pathway formation and selection.

In conclusion, work in the past year extended our knowledge of the molecular rules underlying certain early neural differentiation events. Analysis indicates that we have identified elements of a specific signal transduction mechanism controlling the correct segregation of neuroblasts and involving extracellular signals together with cytoplasmic and nuclear components. Moreover, it seems clear that this interaction mechanism is not exclusively involved in early neurogenesis but is used in various developmental times and tissues and results in the fine tuning of the differentiation of certain tissues, including that of the nervous system.

In addition, the work carried out this year defined new questions and directions. Most significantly, we have started to examine the analogies that may exist between the invertebrate experimental model and vertebrates. This will not only allow for interesting comparisons, but it will also permit us to exploit the experimental advantages offered by the sophisticated vertebrate tissue culture systems. We will be able to address a variety of cell biological questions related to the molecular mechanism within which *Notch* and the other neurogenic genes are integrated. As a first step we have isolated the human *Notch* homologue and found it to be remarkably similar to its invertebrate counterpart.

The Complement System

John P. Atkinson, M.D.—Investigator

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

MANY poorly understood but relatively common and serious human diseases involve aberrations of the immune system.

The function of the immune system is to recognize microorganisms and eliminate them from the host. A major means of accomplishing this is through the synthesis of antibodies and complement components that attach to the foreign substance. These proteins 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. Complement consists of proteins that interact with each other in a reaction resembling a cascade or waterfall: one activates the next. The activated molecules that result from this destroy the invading microbe and promote the inflammatory response. The complement system is a powerful, swift, and highly effective means to fight infection. As might be anticipated, an inherited deficiency of a complement component predisposes an individual to infections.

The production of antibodies is triggered by the invasion of the body by a foreign substance. The binding of antibodies to a target such as a bacterium initiates a series of reactions in which complement proteins swarm onto the surface of the microbe. Such components serve as a ligand for complement receptors on blood cells. A foreign particle soon becomes adherent to and then 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, because 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, the individual makes antibodies to his or her 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, leading 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 viruses 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 it 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. Recently these regulatory proteins have been demonstrated to be expressed in relatively high concentrations 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 means to kill tumor cells or in allowing xenografts to survive in humans.

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 Diversity of Receptor Function

Richard Axel, M.D.—Investigator

Dr. Axel is also Higgins Professor of Biochemistry and 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 research fellowships in Columbia's Institute of Cancer Research, with Sol Spiegelman, 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.

ALL cells communicate with their environment by the interaction of exogenous molecules with receptor molecules on the cell surface. This activates one of many possible intracellular signaling pathways by which information from the environment is decoded to elicit an appropriate response by individual cells or by the organism. Olfaction provides a particularly clear example of the organism's ability to recognize and discriminate a vast array of environmental signals, in this case the molecules perceived as odors.

The cellular immune system provides an example of the organism's ability to respond to diverse foreign substances within the internal environment. These two systems pose the problem of how cells encode sufficient information to recognize and respond to a highly diverse array of signaling molecules. Our laboratory has addressed these problems by analyzing the structure and function of the genes encoding the specific receptor molecules in the immune system, the olfactory system, and the brain.

The Molecular Biology of Smell

Vertebrates have evolved an extremely sensitive mechanism to detect and discriminate a rich diversity of odors. The detection of chemically distinct odors results from the specific association of odorous signaling molecules with discrete receptors on olfactory neurons. These neurons, which reside within the epithelium of the nose, then transmit the information directly to the brain. Thus the detection of odors is restricted to the olfactory neuron, whereas the discrimination of distinct odors requires neural processing in the olfactory bulb and cerebral cortex.

To address the question of how the organism is capable of recognizing more than 10,000 discrete odors, we have recently cloned a large family of genes likely to encode an extremely large number of olfactory receptors. We have isolated a multigene family encoding a vast array of receptor molecules that traverse the membrane seven times. Expression of this gene family is restricted to the olfactory epithelium.

The family comprises more than 200 members, representing one of the largest gene families in the eukaryotic chromosome. All the genes are characterized by sequence and structural motifs shared by all members of the gene family, but the individual genes each exhibit regions of significant sequence divergence. Analysis of the sequences of the members of the family of putative olfactory receptors provides insight into how the family has been generated in evolutionary time and how these genes diverge to accommodate recognition of a large array of odorants.

Isolation of the family of genes encoding the receptor molecules immediately provides one solution to the problem of olfactory perception: How do we recognize so many odors? On one extreme, we could argue that the recognition of diverse odorants could be accomplished by a small number of promiscuous receptors each capable of interaction with several structurally distinct odor molecules. Alternatively, olfactory perception could result from 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.

In a separate series of experiments, we have asked how the association of an odorant with its receptor can generate an action potential transmitted to the brain. Recent evidence from other laboratories suggests that olfactory receptors activate second messenger systems, leading to an elevation in intracellular cyclic nucleotides. We and others have cloned a gene encoding a cyclic, nucleotide-gated ion channel that is unique to

olfactory sensory neurons. This channel traverses the membranes six times, but also contains a domain within the cytoplasm of the cell that binds cyclic nucleotides. The interaction of odors with their specific receptors results in an elevation in cAMP, which in turn leads to the opening of this channel.

The demonstration that the olfactory epithelium expresses a unique ion channel controlled by intracellular levels of cyclic nucleotides now provides a mechanism linking odorant-elicited increases in cyclic nucleotides with the generation of an electrical signal transmitted to the brain. The cloning of a highly diverse multigene family of odor receptors, along with an ion channel responsive to intracellular changes elicited by the receptors, should provide further insight into the recognition mechanisms and neural processing events that permit the discrimination of a diverse array of odors.

Diversity of Recognition in the Cellular Immune System

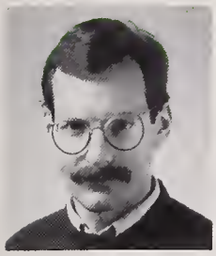
The cellular immune system must recognize and respond to an even greater diversity of signaling molecules in the form of foreign antigens. This diversity is encoded by an extremely large family of T cell receptor genes. Although thymocytes express a T cell antigen receptor, T cells segregate into discrete functional classes. One set, T helper cells, initiate an immune response to foreign antigen by activating other T cells. The second subset, T killer cells, respond to foreign antigen by the elaboration of cytotoxic functions resulting in the death of cells expressing an antigen.

Most T helper cells express an additional surface membrane receptor, the CD4 molecule; most T killer cells express a different but related receptor, the CD8 molecule. We have cloned the genes encoding these two receptors and characterized their functions. Our studies and others indicate that the coordinate engagement of a T cell antigen receptor (TCR) and a CD4 receptor with molecules on the antigen-bearing target cell is required for efficient T cell help. The coordinate engagement of a TCR and CD8 appears to be required for efficient T cell killing.

We have further demonstrated that the T cell surface receptors, CD4 and CD8, are not only required for the mediation of an efficient cellular immune response, but are essential for shaping the repertoire of T cells during development. We have introduced an exogenous CD8 gene into mouse embryos to generate transgenic mice that express a CD8 gene on all T cells. Analysis of the T lymphocytes within the thymus and periphery of these mice leads to a model of T cell development in which the interaction of either CD4 or CD8 with appropriate molecules on the surface of cells within the thymic epithelium is a primary determinant of whether a cell will be either a helper or a killer cell.

Thus our studies on the structure and function of receptors from the immune system, olfactory system, and brain provide examples of how the eukaryotic chromosome can generate the diversity required to recognize and respond to a vast array of signaling molecules in the environment.

Developmental Genetics in the Mouse and Human



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-UCLA Hospital and 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 abnormalities such as miscarriages and birth defects.

In organisms traditionally subject to experimental genetic analysis, like 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. My laboratory is studying 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 transgenic mice.

Genomic and Functional Characterization of the Mouse Segmentation Gene *kreisler* (*kr*)

The mouse *kr* gene, originally recognized by its effects on inner ear development and craniofacial morphogenesis, interferes with the normal formation of metameric units in the developing hindbrain, so-called rhombomeres. The *kr* mutation is x-ray induced and therefore likely to represent a structural rearrangement that alters a relatively large region of the chromosome. Located within two map units of *kr* are two additional loci that affect fundamental aspects of peri-implantation development: *brachypod* (*bp*), which produces limb reduction abnormalities similar to those seen in the inherited human disease Osebold-Remondini syndrome, and *agouti* (*A*), which represents at least three mutations that, when homozygous, are lethal at or around the time of implantation.

Toward the eventual goal of isolation of these genes, a physical map of mouse chromosome 2 surrounding *kr*, *A*, and *bp* has been constructed with a cell fusion technique called radiation hybrid mapping. This technique, developed by David Cox, is based on the likelihood that pieces of mouse DNA closely linked to each other will stay together, or "cosegregate," after x-irradiation and fusion to hamster cells. Our results have led to the isolation of a mouse/hamster hybrid cell line that contains the *kr*, *A*, and *bp* loci along with very little "extra" mouse DNA. One of the mouse DNA fragments present in this hybrid cell line, closely linked to the mouse *Src* gene, detects a chromosomal alteration in *kr/kr* mice. The alteration involves a chromosomal break located approximately 100,000 base pairs away from the beginning of the *Src* gene, but does not interrupt the *Src* coding sequences. Small DNA fragments recovered and mapped by this technique are now being used to reach the *kr* gene by the serial isolation of overlapping cosmid and yeast artificial chromosome clones.

To characterize further the relationship between rhombomere formation and *kr*, we have studied the expression of a group of genes, the *Hox-2* cluster, thought to play a role in determining the identity of individual rhombomeres. Different members of the cluster normally exhibit anterior borders of expression in adjacent rhombomeres. In collaboration with Michael Frohman and Gail Martin, we have shown that, in *kr/kr* embryos, several *Hox-2* family members are expressed in the "wrong" rhombomere(s), which suggests that the *kr* gene product may affect neural tube segmentation by controlling the expression of *Hox-2* genes. By isolating the *kr* gene and further characterizing the *kr/kr* phenotype, we hope to learn more about the molecular pathways of mammalian segmentation and the role of *Hox-2* genes in rhombomere morphogenesis.

Conditional Inactivation of Recessive Genes in Embryonic Stem Cells

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 antisense 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.

Cell Cycle Control

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 Sussex and with Amar Klar at Cold Spring Harbor Laboratory.

THE division of one cell into two, generally known as the cell division cycle, is among the most basic of biological processes. The underlying molecular pathways that regulate and effect the cell cycle are gradually being elucidated by a combination of genetic, biochemical, and cell biological studies.

The two key events in the cell cycle are the replication of nuclear DNA, which occurs during the S phase, and cell division or mitosis, often referred to as M phase. Genetic studies, using the fission yeast, a single-celled organism, have revealed that both the S and M phases of the cell cycle are regulated by a single enzyme. This is a protein kinase (which transfers phosphate from ATP to a target substrate, altering the substrate's biological properties) of which the catalytic subunit is called cdc2. This enzyme is not only present in yeast cells, but also in those of higher vertebrates, including humans.

The cdc2 acts as a multifunctional enzyme by virtue of its association with a class of proteins called cyclins. Each cyclin, of which many types have been identified, binds to cdc2 and confers on the protein kinase a particular substrate specificity and also targets it to particular cellular locations. Different cyclins are synthesized and degraded at particular stages of the cell cycle, thereby creating different patterns of cdc2 activity during the S and M phases of the division cycle.

The critical cdc2 kinase is not regulated exclusively by passive association with cyclins, but also by its own state of phosphorylation. In particular, cdc2 is one of the most heavily tyrosine-phosphorylated proteins in actively proliferating cells. Tyrosine phosphorylation serves to inactivate the M-phase form of cdc2 (cdc2/cyclin B), and dephosphorylation of the enzyme triggers mitosis.

Our current research is focused on the phenomena discussed above. In particular, a problem of present interest is the mechanism that coordinates the different phases of the cell cycle. For example, why does cell division not occur until DNA replication is completed? The genes that control such cellular decisions are being investigated.

*Fission yeast cells have been stained with a dye (DAP1) that reveals DNA and hence shows the cell nucleus. These particular mutant cells lack two genes (*mik*1 and* wee*1) that encode the inhibitory cdc2 tyrosine kinase. They are therefore entering mitosis prematurely, before the completion of DNA replication, and are dividing themselves to death.*

Research of David H. Beach.

Homeotic Gene Action 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.

THE most prevalent developmental strategy in the animal kingdom is segmentation—that is, division of the embryo into a series of similar segments that later differentiate and become specialized for particular functions. The underlying molecular events are best understood in the fruit fly *Drosophila*, where segmentation and segmental differentiation are governed by genes expressed in a temporally and spatially ordered hierarchy. Many of the 50-some genes within this hierarchy have been characterized at the molecular level; most encode proteins implicated in the control of gene expression. Much progress has been made in learning how these regulatory genes and their products interact to specify each other's expression and thereby generate a detailed system of spatial information.

Little is known, however, about how this spatial information (in the form of localized regulatory proteins) is used to assemble precisely the complex pattern of structures that arise in the course of embryogenesis. My laboratory's long-term goal is to extend knowledge of this genetic hierarchy by identifying and characterizing target genes whose products are more directly involved in the construction of the embryo than are proteins within the regulatory hierarchy. We are also studying the mechanisms by which such target genes are differentially regulated.

The first tiers of the regulatory hierarchy contain maternally expressed genes, whose products establish broad polarities within the embryo. The segmentation genes of the middle tiers subdivide the embryo along the anterior/posterior axis into a linear array of homologous segments. The eight homeotic genes of the final tier are regulated by the segmentation genes and function to produce the specialized structures that distinguish the segments.

The basic morphogenetic pathways for the repeated segmental unit appear to be controlled, at least in part, by regulatory input from segmentation genes. Homeotic genes, however, are the laboratory's current focus because of their unique roles in channeling the processes of morphogenesis into pathways that produce the diversity of structures distinguishing the segments. The proteins encoded by homeotic genes each contain the homeodomain, a 61–amino acid segment associated with DNA-binding activity that is also present in some of the segmentation and polarity genes. Homeodomain proteins have sequence-specific DNA-binding properties and are generally involved in control of gene expression at the level of transcription.

The homeodomain, first identified in *Drosophila*, has been found in all other multicellular animals examined, including vertebrates. Perhaps the most striking aspect of this evolutionary conservation is that in vertebrates, homeodomain genes closely related to the *Drosophila* homeotic genes (as judged from conservation of homeodomain sequences) are clustered in the same linear order found in the *Drosophila* homeotic gene complexes. In addition, expression along the body axis is colinear, with locations of particular genes within the vertebrate clusters in a manner strikingly similar to that in *Drosophila*. These observations suggest that some of the mechanisms of positional specification, and perhaps even some aspects of segmentation and segmental differentiation, are conserved between insects and vertebrates.

Differential Recognition of DNA Sequence by Homeotic Proteins

A major determinant of a regulatory protein's effect upon a target gene is the affinity of that protein's interactions with nearby DNA sequences. Therefore a key component of any transcriptional regulator's biological properties is the regulator's sequence preferences in binding to DNA. The proteins encoded by *Drosophila* homeotic genes pose something of a puzzle in this regard because, despite great sequence similarities between their homeodomains, they all have distinct biological functions (i.e., they each implement distinct segmental differentiation pathways). To enable systematic evaluation of DNA sequence preferences, we have developed a method by which the optimal DNA-binding sequence for a particular protein can be selected

from a large population of molecules that includes all possible sequence combinations.

This method has been used to determine and compare the optimal binding sites for two closely related homeodomain peptides encoded by the *Drosophila* homeotic genes *Ultrabithorax* (*Ubx*) and *Deformed* (*Dfd*). The optimal sites for these proteins hold in common a core sequence, TAAT, that is very important for binding. Additional bases to either side of the TAAT core are also recognized, but these base preferences differ between the two proteins. Comparative studies indicate 5- to 10-fold relative preferences of *Dfd* and *Ubx* homeodomains for their own optimal sites. Additional studies are under way to determine which of the amino acid sequence differences between *Ubx* and *Dfd* are responsible for the differences in DNA recognition. Additional insight into the nature of the critical molecular interactions should come from a crystallographic study in collaboration with Carl Pabo (HHMI) and Neil Clarke, at the Johns Hopkins University, of a *Ubx* homeodomain complex with its optimal binding site.

In vivo studies with the full-length proteins indicate that their DNA sequence preferences are similar to those of the homeodomain peptides. The full-length *Ubx* protein, however, unlike the *Ubx* homeodomain peptide, binds cooperatively to multiple sites. This cooperativity can increase complex stability from less than a minute for a single isolated site to several hours for four or more sites.

The most striking property of these cooperative interactions is that they can occur between sites separated by a variety of distances, ranging from a single helical turn of the DNA to more than 20 turns. The long-range interactions involve a mechanism in which the intervening DNA is curved to permit contact between proteins bound at the base of a resulting DNA loop. The affinity for binding of full-length *Ubx* protein is therefore a function of the number of sites present within a DNA fragment, and *Ubx* proteins appear to act as integrators of the number and affinity of core binding site sequences.

With regard to differential gene regulation, these results suggest the following: first, differences in DNA sequence recognition account for at least some of the biological specificity of homeotic genes; second, the molecular basis for discrimination between homeotic protein–binding sites may be interactions with bases flanking a common TAAT core; third, large differential affinities for a DNA region might be generated by cooperative binding of homeotic proteins to collections of core sites that individually show only small differences in affinity.

Identification of *Ubx* Regulatory Targets

We have used a method similar to the one for determination of optimal DNA binding sites to select and clone DNA fragments containing *Ubx* binding sites from total *Drosophila* genomic DNA. We expect that genes regulated by *Ubx* will be located near such binding sites. We have also identified sequences from *Drosophila* cultured cell lines that are differentially expressed in the presence and absence of *Ubx* protein. The sequences isolated by these procedures are being analyzed to determine how they function in the embryo.

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 the NIH.

RESEARCH in our laboratory is aimed at molecular analysis of human genetic diseases, with the dual goals of learning more about the diseases themselves and shedding light on normal human biology. Currently we are studying cystic fibrosis, deficiencies in urea cycle enzymes, spinocerebellar ataxia, and defects in cell adhesion molecules. Most of this research is conducted in close collaboration with William O'Brien.

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. Since 1985 our laboratory has used DNA analysis for carrier detection and prenatal diagnosis for CF. This work was changed dramatically in late 1989 when other laboratories cloned the gene for CF and identified the most common defect (mutation) in the gene. We analyzed more than 200 North American families for this mutation; the defect was present on 76 percent of the CF chromosomes. This defect was present on only 30 percent of CF chromosomes in Ashkenazic Jewish families, however, indicating that other mutations predominate in that population. Direct analysis of mutations greatly improves the genetic diagnostic tests that can be offered to families.

One option for prevention of CF would be to test all couples prior to reproduction, to identify carriers of the abnormal gene. Although a single common defect is present in the majority of CF chromosomes, dozens of different mutations affect the remaining fraction of abnormal chromosomes. Thus no one simple DNA test can be used to identify all gene carriers. Our laboratory and other laboratories participating in a large collaboration have identified many of these additional mutations. The laboratory has developed more efficient methods for detection of multiple mutations. We currently test for the five most common mutations causing CF in North Americans, and this identifies 85 percent of carriers. The ability to detect 90–95 percent of carriers would be preferable before offering carrier testing to all reproductive-age couples, and the laboratory is working to achieve this level of detection.

To develop a treatment for CF, it is desirable to have an animal model for easier analysis and for therapeutic trials. Methods are available to develop an animal model in mice, and we have characterized the sequence of the mouse CF gene as one step in this process. The mouse gene, which is similar to the human gene, has been used to prepare DNA clones that are suitable for disrupting the normal mouse gene in cultured cells. These altered cells can be used to generate mice affected with CF. Despite substantial effort, it has proven difficult to disrupt the CF gene in the mouse. It would also be desirable to study the equivalent of the CF gene in *Drosophila* (fruit flies) or yeast, if such a gene exists. Preliminary work has identified some related genes in these organisms, but it is not clear if these genes are functionally equivalent to the human CF gene.

For many years our laboratory has studied enzymes of the urea cycle, with particular focus on argininosuccinate synthetase (AS). Genetic deficiency of AS causes citrullinemia, a disease characterized by accumulation of ammonia in the body, which leads to mental retardation, coma, and death. Patients with citrullinemia have a recessive disease, with two defective copies of the gene. Mutations causing citrullinemia include deletions of segments of the gene, numerous different single–amino acid substitutions, and defects that cause abnormal splicing of the messenger RNA. Fourteen different mutations have been defined; almost all patients inherit a different mutation from each parent. The laboratory has also used citrullinemia as a model disease for developing somatic gene therapy. The human gene for AS was introduced into retrovirus vectors in the form of a cDNA. The viruses produced are able to infect mouse and human cells, and they have been used to infect mouse bone marrow cells, which are then transplanted into lethally irradiated recipient mice. The human gene for AS is expressed for many months in a proportion of the transplanted animals, and this long-term expression in mice is encouraging.

In collaboration with Huda Zoghbi, we are attempting to clone the gene for dominantly inherited spinocerebellar ataxia, a neurological disease that is passed from generation to generation. The symptoms usually begin in young adult life and result in physical disability, mental deterioration, and death within about 10 years. Recent studies have localized the gene to a small region of chromosome 6. Large DNA clones known as yeast artificial chromosomes have been isolated from this region, and the tedious task of identifying and cloning the disease gene is continuing.

A major new project is the study of cell adhesion molecules, which are involved in cell-surface and cell-cell adherence, particularly as they relate to interactions between white blood cells and the walls of blood vessels. The genetic disease leukocyte adhesion deficiency involves the absence of one of these molecules, designated CD18, on the white blood cells and results in fatal susceptibility to infections. Some mutations causing this disease have been determined. The disease is an excellent model for attempts at somatic gene therapy, and the gene has been introduced into retroviral vectors. A major focus of this work is to generate mouse models with mutations in these cell adhesion genes. The mouse genes for CD18, intercellular adhesion molecule-1 (ICAM-1), and granule membrane protein-140 (GMP-140) have been isolated and sequenced. For CD18 and ICAM-1, the mouse gene has been disrupted in cultured cells, and the cultured cells are being injected into mouse embryos to obtain mutant animals. Mouse mutants would be valuable for studying the normal biologic function of the various cell adhesion molecules. These molecules are likely to be important in processes such as infection, inflammation, and atherosclerosis.

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.Sc. degree in zoology and M.Sc. 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 and the Rolf Luft Award from the Swedish Medical Society.

DIABETES mellitus is a disorder of carbohydrate metabolism characterized by elevated blood glucose levels. It affects individuals of all ethnic groups. In the United States, an estimated 6 million persons are known to have diabetes, and there is 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, which results from reduced insulin levels in some patients and a relative deficiency in others, 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 such as atherosclerosis and hypertension, genetic factors contribute to the development of NIDDM. Our long-term goal is to identify the genes for diabetes susceptibility 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 applies 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 expression of the major susceptibility genes.

In our genetic studies we are examining the segregation of DNA polymorphisms in diabetes-prone families. We have recently identified a DNA marker on the long arm of human chromosome 20 that cosegregates with NIDDM in a larger multigenerational family with a form of NIDDM termed maturity-onset diabetes of the young (MODY). The diabetes in this family is characterized by low serum insulin levels, an age of onset of less than 25 years, and an autosomal dominant mode of inheritance. The isolation of this MODY-susceptibility gene and identification of the mutation that impairs its function may provide important clues to causes of other forms of NIDDM. In addition, it will also be possible to assess the contribution of this locus to the overall prevalence of diabetes mellitus. Since the MODY-susceptibility gene on chromosome 20 is unlikely to account for genetic susceptibility to NIDDM in all individuals, we are continuing to search for other markers that cosegregate with NIDDM, using large families in which the diabetes is not linked to the marker on chromosome 20. These genetic studies should lead to the identification of additional diabetes-susceptibility genes.

Drawing on our understanding of the pathophysiology of NIDDM, we are also cloning and characterizing genes that might reasonably contribute to diabetes susceptibility. These include genes encoding proteins that are expressed in the insulin-producing cell, including insulin, islet amyloid polypeptide (a newly discovered hormone-like peptide of uncertain function), and the K^+ and Ca^{2+} channels that are involved in insulin secretion. We are also studying the genes for proteins responsible for uptake and metabolism of glucose, since the insulin-producing cells of NIDDM patients have lost the ability to secrete insulin in response to increasing circulating glucose levels. Other candidate genes encode proteins expressed in the insulin-responsive target cell, including the insulin receptor and the membrane proteins involved in glucose uptake. In addition, we have considered genes encoding serum proteins that are instrumental in the trans-

port of triglycerides and cholesterol and thus implicated in the development of cardiovascular disease, since there are data suggesting that diabetes and cardiovascular disease may have common genetic antecedents.

Genetic variation associated with these candidate genes is being examined in groups of diabetic and nondiabetic subjects, as well as in families, to assess the contribution of these genes to the development of NIDDM. The results of the studies to date suggest that there is variation in the genes for the insulin receptor and for apolipoproteins A1 and B that influences the phenotypic expression of a major susceptibility gene.

Our isolation of the genes for membrane proteins involved in the transport of glucose across the plasma membrane, in addition to providing markers for our genetic studies, has 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.

Data from our laboratory and others indicate that decreased levels of two of these glucose transport proteins may contribute to the development of the diabetic state. In adipose tissue there are reduced amounts of the glucose transport protein GLUT4, which is responsible for insulin-stimulated glucose uptake by this tissue. The reduced amounts of GLUT4 provide a molecular explanation for the decreased ability of adipose tissue of diabetic and obese individuals to take up glucose. Because adipose tissue participates in the disposal of dietary glucose after a meal, the decreased expression of this protein may contribute to high circulating glucose levels. The levels of GLUT2, the glucose transport protein of insulin-producing cells, are also decreased in diabetes. Since GLUT2 may comprise part of the glucose-sensing mechanism that regulates insulin secretion, decreased levels of this protein may be responsible for the abnormal functioning of the insulin-producing cells of diabetic individuals. A better understanding of the regulation of expression of GLUT2 and GLUT4 may facilitate the development of novel therapeutic strategies for the treatment of diabetes.

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, Division of Neuroscience, and Department of Cell Biology, 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*. We believe that the results will help us understand the development of the nervous system in 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 two genes that are expressed at the outset of embryonic nervous system development and later in embryogenesis in most cells of the PNS. The genes were identified in enhancer detector screens in which regulatory sequences are determined by means of a β-galactosidase reporter gene. The newfound genes are essential, since mutations cause embryonic lethality. Here we describe their preliminary characterization.

One of the two genes is named *couch potato* because some homozygous mutant flies are viable but hypoactive and have poor jump responses, poor flight abilities, and slow recovery from ether anesthesia. Some of these attributes are probably due to thoracic muscle defects that we observe in the adult fly. More than 10 insertional mutations in *couch potato* were recovered from several enhancer detector screens, and a careful genetic analysis revealed that none 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.

In order to define the null phenotype, we performed an excision mutagenesis of one of the insertions and recovered an embryonic lethal mutation in the gene that, by genetic criteria, can be defined as a complete loss-of-function mutation. Some embryos lacking the *couch potato* gene exhibit defects in the neuronal pathways of PNS and lack some peripheral nerve cells. Hence the *couch potato* gene seems to play a role in the development of the PNS as well as in the development of the adult thoracic muscles.

A molecular analysis of *couch potato* was initiated with a genomic DNA fragment next to one of the enhancer detector insertions. This fragment allowed us to isolate three different *couch potato* cDNAs. These were labeled for a nonradioactive whole-mount *in situ* hybridization technique, which showed that the gene is expressed in the PNS, including the glial cells, and in some glial cells of the central nervous system. In all cases the expression patterns of β-galactosidase in the embryos that carry an enhancer detector and the transcripts of *couch potato* are essentially the same.

All the insertions that were recovered are clustered in a 450-bp genomic fragment. In addition, there is a very good correlation between the defects that we observe in embryos or adults and the position of these insertions; insertional alleles that cause a particular phenotype are clustered within 50 bp. These data, combined with the genetic data, indicate that the genomic sequences next to the insertions contain some of the gene's regulatory sequences that confer the tissue-specific expression pattern. That this is indeed the case has been shown by bringing these sequences upstream of a β-galactosidase reporter gene in transformed flies. The transformed embryos express the reporter in a subset of cells (PNS of head and glial cells) in which *couch potato* is expressed, showing that these sequences contain at least some *couch potato* regulatory sequences.

The expression of the *couch potato* gene in the glial cells is interesting, as we observe defects in the pathways of the peripheral neurons in a substantial fraction of the embryos lacking the gene. The glial cells may play a key role in helping the neuron find its normal target organ, tissue, or cell. It is therefore possible that the lack of expression of the gene in the glial cells causes a neuronal pathway defect.

To test the hypothesis that the glial cells play a role in neuronal pathway formation, we would like to ablate them. Hence we have devised a technique that should allow specific ablation of these cells early in development. It makes use of a tem-

perature-sensitive diphtheria toxin, DT-Ats, that we have developed in yeast and tested in *Drosophila* eye cells.

Sequence analysis of the different *couch potato* cDNAs indicates that the gene may have peculiar features. The three cDNAs share a long open reading frame that conforms with the *Drosophila* codon usage, but we have been unable to define an initiation codon. Translation of the open reading frame indicates that the gene may encode a transmembrane protein that shows no homology to any of those published in the data banks.

In summary, we believe that *couch potato* may play an important role in the development of the PNS, as its gene product may help guide the axons of the PNS neurons to their appropriate target. The role of the gene in adult thoracic muscle development remains elusive, but many phenotypic similarities have been observed between *couch potato* and *stripe*, which maps close to the former gene. The *stripe* gene also affects thoracic muscle development and is thought to be required in the nervous system.

The second gene we are studying, named *A37*, is expressed early in the development of the PNS. It maps at cytological band 80A and, like *couch potato*, it was first detected in an enhancer screen. So far, only a single insertion has been identified, to our knowledge. Flies homozygous for this particular enhancer detector insertion are homozygous viable. A mutagenesis experiment allowed us to recover 10 embryonic recessive lethal mutations that fail to complement one another. The phenotype of the mutant embryos is being characterized, and preliminary data suggest that some cells of the embryonic PNS are lacking, although the gene is expressed in almost all PNS cells, including the support cells. The gene is apparently encoded in a 4.6-kb transcript, which is only expressed transiently in the PNS of developing embryos (5–10 h). Later in embryonic development (14–16 h), the transcript is also observed in most embryonic muscles. Preliminary sequencing data show that the cDNAs that we have isolated so far contain no open reading frames of significant length. We are presently trying to define the phenotype of the mutants more accurately and continuing to characterize the gene at the molecular level.

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.

THE production of mature blood cells involves an ordered series of differentiation programs. At the root of this production system are the pluripotent hematopoietic stem cells—a small population of cells capable of extensive proliferation and differentiation, but also of self-renewal. These cells, which normally reside in the bone marrow, arise in early fetal development and persist throughout adult life. If removed from the bone marrow and transplanted into a prepared recipient, they will stably reconstitute the full blood and immune cell developmental systems. We are working out techniques for efficient gene transfer into mouse and human stem cells, using retrovirus vectors. Ultimately we hope to exploit these methods for human gene therapy of genetic and acquired diseases.

A viral vector system based on the Moloney murine leukemia virus (MLV) has been chosen because of its theoretical potential for high gene transfer efficiency in a variety of mammalian cell types (including human). 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. MLV vector particles are able to carry their genetic material to the target cells, but are unable to replicate and spread as a live infectious agent.

We have been studying two model systems that allow investigation of several fundamental properties of the stem cells as targets for gene transfer and for expression of genetic material by the vector. One model uses the bacterial antibiotic resistance gene *neo* to introduce identifiable genetic tags into individual stem cells. The other 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 retrovirus 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 allow a smooth transition to 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 during exposure to the vectors.

In collaboration with Doug Williams (Immunex, Seattle), we have been evaluating the effects of several recombinant hematopoietic growth factors on retroviral vector–mediated gene transfer into stem cells. These factors have 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.

We observed that LIF stimulated a 10-fold increase in the efficiency of gene transfer into the primitive hematopoietic precursor cells called CFU-S (colony-forming unit, spleen). A novel assay employing inbred transgenic mice was then 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 now allows about 70 percent of the stem cells to be infected with the vector. Subsequently all the mice receiving such cells maintain high-level expression of human ADA in their blood and immune system organs for at least six months. Suspecting that LIF acts in concert with one or more other growth factors in our experimental model, we are examining its effects alone and in combination with other stimulators on purified hematopoietic stem cells.

To improve analysis of the behavior of stem cells and their progeny in culture and after transplant, we have developed a new method for introducing unique genetic identifiers into individual cells. This biochemical method, which

allows identification of the stem cells by the polymerase chain reaction (PCR), 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, variations conveniently distinguished by PCR. When mixed together these vectors provide an array of potential markers that can integrate into each stem cell. Because each cell receives one to four vectors, the integration of a particular subset of the vectors provides a genetic fingerprint unique to each cell family.

Preliminary results with these vectors indicate that we can mark the most primitive hematopoietic precursors efficiently. We plan to use this method to compare different culture conditions for stem cells. We hope to be able eventually to grow the stem cells in very long-term cultures.

The retroviral vector genetic marking system may also prove to be highly informative in clinical bone marrow transplantation. We have undertaken collaborative studies with Albert Deisseroth (M.D. Anderson Hospital, Houston) to determine whether the techniques that have been used to optimize gene transfer of mouse bone marrow stem cells can be used for human cells.

These experiments demonstrate that consistent and efficient gene transfer into primitive human precursors is possible. After further tissue culture studies are performed, we plan to do genetic labeling experiments in patients undergoing bone marrow transplant for advanced leukemia. These investigations will provide a novel way to investigate the origins of leukemia recurrence and the recuperation of the blood system following transplantation. We also hope to learn many valuable lessons that will contribute to the long-term goal of treating disease by insertion of therapeutic vectors.

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 School of Medicine. His postdoctoral training in membrane protein biochemistry was completed at Harvard University with Daniel Branton. Before joining Duke University, Dr. Bennett was a member of 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 activities of cells, such as cell motility, organization of the cytoplasm, 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.

Ankyrin in the Nervous System

Brain ankyrin binds to brain spectrin and to integral membrane protein sites in brain membranes. Ankyrin appears to function as an adapter between these proteins and the spectrin skeleton. We recently discovered that brain contains multiple forms of ankyrin, which all bind to spectrin but are likely to associate with distinct membrane proteins.

One well-characterized ankyrin-binding protein is the voltage-dependent sodium channel. An 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. The localization of the voltage-dependent sodium channel at nodes of Ranvier is important for normal conduction of nerve impulses. We plan to explore the possibility that linkage of the sodium channel to ankyrin plays a role in either confining or initial targeting of the sodium channel to the nodes of Ranvier. These studies are relevant 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.

We recently determined the complete amino acid sequence of the major form of ankyrin in human brain. The gene encoding brain ankyrin is located on chromosome 4; the gene encoding erythrocyte ankyrin is on chromosome 8. Surprisingly, the same gene that encodes erythrocyte ankyrin also is expressed in brain, with high abun-

dance in the cerebellum and brain stem. Brain thus contains ankyrins encoded by different genes. The erythrocyte form of ankyrin is missing in a strain of mutant mice developed at the Jackson Laboratory. The ankyrin-deficient mice experience degeneration of Purkinje cells, a major type of neuron in the cerebellum, and develop a stagger and difficulty in walking. These mice retain ankyrin at the nodes of Ranvier, suggesting that ankyrin at the nodes is the product of yet another gene. The neurological problems of ankyrin-deficient mice may have counterparts in humans with slowly progressive diseases due to death of nerve cells.

The diversity of ankyrins suggests that this family of proteins may interact with many membrane proteins in addition to the sodium channel. We have used the membrane-binding domain of the major form of brain ankyrin to isolate ankyrin-binding proteins and have identified at least 10 new proteins. The consequence of linkage of these proteins to the membrane skeleton could be to localize them in specialized regions of the cell membrane appropriate for their function.

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 recently 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. We are currently characterizing a protein that is a candidate for a role in linking spectrin to the membrane.

Ankyrin Structure

Ankyrin contains three independently folded domains: one domain interacts with certain membrane proteins, another associates with spectrin, and a third 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 recently discovered that this conserved portion of the ankyrin sequence is responsible for 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 and expect that this structure will increase understanding of the 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. 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.—Assistant 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 his present position.

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. These cells, originally derived from a white blood cell, the monocyte, exist in many tissues.

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. One, it causes increased transcription of the TNF gene, leading to a marked accumulation of TNF mRNA within macrophages. Two, 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 cells that produce TNF, in other words, CAT synthesis also occurs. 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. While the protein does not appear to be produced elsewhere in healthy animals, it is readily induced by administration of lipopolysaccharide (LPS) or by various authentic infections.

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 blocking 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 dif-

ferent locus from that of glucocorticoids. We have shown that the two classes of drugs when combined assert 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 through investigation of the many effects of TNF in health and disease and may also be useful as a therapeutic tool.

T Lymphocyte Ontogeny and Specificity

Michael J. Bevan, Ph.D.—Investigator

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

T lymphocytes mature in the thymus from bone marrow–derived hematopoietic stem cells. A key part of that process is the rearrangement of T cell receptor genes, the products of which will be expressed on the surface to recognize foreign antigens. The receptors, while highly varied, are expressed clonally: each T cell expresses only one kind of receptor.

Two functionally different types of T lymphocytes emerge from the thymus to populate the peripheral lymphoid organs. These are usually called helper and cytotoxic. In response to recognizing foreign antigen, helper T lymphocytes release lymphokines, which augment the response of other lymphocytes, including B cells, to the antigen. Cytotoxic T lymphocytes specifically recognize cells that have become infected with intracellular pathogens and kill them efficiently.

Although the same pool of receptor genes is used by both types of T lymphocytes, they are able to respond to quite different types of antigen by virtue of the different classes of MHC (major histocompatibility complex) molecules engaged in antigen presentation. The MHC encodes two types of surface glycoproteins: class I and class II. Killer cells recognize peptide components of foreign proteins bound in the groove of class I MHC molecules, while helper cells recognize peptide antigens presented in the groove of class II molecules.

Class I MHC glycoproteins are assembled with β_2-microglobulin and a foreign peptide fragment shortly after synthesis in the endoplasmic reticulum. The peptides that become part of the class I complex are derived from intracellular self and foreign (viral) proteins. Class I MHC molecules do not associate with antigens that are taken up from the extracellular fluid. This is the world of class II presentation. Soluble extracellular antigens endocytosed by antigen-presenting cells are digested, then bound by class II molecules in endosomes. Because of this dichotomy in antigen presentation, the body mounts a helper T cell response to noninfectious extracellular antigens, and it mounts a cytotoxic T cell response to cells infected with viruses or bacteria.

The generation of appropriate T cell receptors by gene rearrangement is not left to chance. Immature T cells expressing new receptors on their surface are only allowed to mature if the receptor has an appropriate affinity for class I or class II self MHC molecules. This process, referred to as "positive selection of the T cell receptor repertoire," is quite stringent. Ninety percent of immature T cells expressing new T cell receptors fail this selection step and die in the thymus.

Selecting the T Cell Repertoire

T cells recognize fragments of foreign protein antigens bound in the groove of MHC-encoded molecules. The peptide-binding groove of the MHC molecule is extremely polymorphic, so that different alleles bind a different range of peptides. We know the crystal structure of a class I MHC molecule, and we have a good guess as to how peptide lies within it. Although we do not know the structure of the T cell receptor, a number of groups have suggested that the receptor recognizes self MHC plus foreign peptide by contacting their residues on the face of the complex.

Different alleles of MHC molecules expressed in the thymus select quite different T cells for maturation based on their receptor specificity. Since positive selection of the T cell repertoire is done in the absence of foreign peptides, it was commonly supposed that positive selection would be based on the affinity of the T cell receptor for the exposed outer faces of the MHC. We were able to show, however, by comparing the T cell receptor repertoire selected by variant class I molecules, that residue changes within the peptide-binding groove—i.e., not exposed on the outside of the molecule—determine the specificity of selection. The immediate interpretation is that the MHC class I groove is occupied by a self peptide during positive selection in the thymus, and the type of peptide is all important for the result. Our studies also suggest that a large variety of self peptides presented by the same class I molecule select for different portions of the T cell repertoire. These unknown self peptides are

therefore set up as mimics of the foreign antigenic world, predicting the antigenic universe that the immune system will encounter after thymic maturation.

Class I MHC Antigen Presentation

The class I heavy and light chains have amino-terminal signal sequences so that they are made on membrane-bound ribosomes and assemble in the endoplasmic reticulum. Many of the peptide antigens that they complex with and present to cytotoxic T cells, however, derive from internal viral proteins that are made in the cytoplasm on free ribosomes. It seems likely that there is a peptide transporter located in a pre-Golgi membrane that allows the transfer of unfolded molecules or peptide fragments from the cytoplasm into the compartment occupied by nascent class I molecules. Because of the nature of class I–restricted antigen presentation, the readily and cheaply available protein antigens such as egg proteins and serum proteins could not be used to stimulate a cytotoxic response. We overcame this problem by using the genes for chicken ovalbumin and β-galactosidase transfected into mouse cells to generate such class I–restricted responses. By this entirely artificial procedure, we converted these antigens into endogenously synthesized mouse proteins and were readily able to stimulate a class I–restricted cytotoxic response.

This system has been used to study how antigen gains access to the class I pathway of antigen processing and presentation. Soluble protein antigen can be introduced into the cytoplasm and then into the class I antigen-processing pathway by the osmotic lysis of pinosomes or by the uptake of acid-sensitive liposomes that fuse with endosomal membranes. We have used these systems to study the kinetics of antigen breakdown and presentation and to ask what degradation systems are involved in antigen processing for class I–restricted presentation.

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 then a Ph.D. degree in biochemistry and molecular biology from Harvard University, where her thesis advisor was Don Wiley. She completed a crystal structure analysis 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.

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.

My previous work in the laboratory of Don Wiley (HHMI, Harvard University) used x-ray crystallography to determine the structure of a histocompatibility molecule. We learned how and where viral molecules bind to histocompatibility protein and how T cell receptors might bind to this complex. 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 will also try to make crystals of a complex between a T cell receptor and histocompatibility molecule to be used in a three-dimensional structure determination. 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 the 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 these proteins than would be otherwise possible. We have expressed fragments of histocompatibility molecules in bacteria in preparation for making complexes between these proteins and antigenic peptides. We are also working on protein expression in mammalian cells.

In collaboration with Kai Zinn at Caltech, we have expressed in a hamster fibroblast cell line large quantities of a cell surface protein found in the grasshopper and fruit fly nervous systems. We are interested in the structure of this molecule, because many molecules involved in cell surface recognition in the immune and nervous systems are related to each other. This protein has now been purified and crystallized. We are continuing to investigate ways to improve the quality of the crystals, as the current crystals do not diffract to sufficient resolution to allow a crystal structure determination.

We have also used a similar protein expression system to make an Fc receptor, a protein that binds to the Fc portion of antibody molecules. (This work is in collaboration with Neil Simister, Brandeis University.) This particular Fc receptor is found in the intestine of newborn mammals and binds maternal immunoglobulin in 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 are currently working out a purification scheme for this protein and will try to crystallize the protein in preparation for an x-ray structure determination. We also have crystals of a structurally unrelated Fc receptor and will compare the binding of these two receptors to their common ligand.

Molecular Mechanisms of Insulin and Polypeptide Hormone Action

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

Dr. Blackshear is also Professor of Medicine and Assistant Professor of Biochemistry at Duke University. He received his D.Phil. degree in biochemistry at Trinity College, Oxford University, and his M.D. degree from Harvard Medical School. Before moving to Duke University, Dr. Blackshear 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 has been studying the molecular mechanisms of action of insulin and other polypeptide growth factors. The goal of these studies is to understand the biochemical steps between the binding of insulin to its receptors on the surface of its target cells and the ultimate stimulation of changes in enzyme activity, protein synthesis, and gene expression. An understanding of the molecular details of insulin's actions may lead to the biochemical characterization of the insulin resistance seen in certain common disorders, such as obesity and type II diabetes, and possibly to novel therapies for these disorders.

We are studying the intracellular actions of insulin and related polypeptide growth factors at several levels. One of the first effects of insulin on cells is the activation of an enzyme activity intrinsic to those receptors, a protein kinase activity that transfers a phosphate group from the cellular energy source ATP to a specific amino acid, tyrosine, in target proteins. Later effects of insulin include the activation of other cellular protein kinases that catalyze the phosphorylation of various cellular proteins on serine and threonine residues. In the past year we identified one of these insulin-activated kinases as the Raf-1 protein kinase, a proto-oncogene product that is related to the transforming oncogene of a murine sarcoma virus. The activation of this protein kinase was accompanied by its phosphorylation on serine and threonine residues. In our laboratory Ruey-Min Lee recently identified two insulin-activated protein kinases that in turn phosphorylate the Raf-1 protein; these kinases are activated within two minutes of cellular exposure to insulin and may represent one of the missing links between the tyrosine kinase activity of the insulin receptor and the serine kinase activity of the Raf-1 protein kinase.

Another well-known action of insulin is the stimulation of protein synthesis in its target cells. The synthesis of many proteins is increased as part of insulin's general anabolic effect. However, insulin also stimulates the synthesis of several proteins to a much greater extent than the overall average of cellular proteins; the synthesis of many of these proteins appears to be stimulated at the level of messenger RNA translation. Joyce Manzella is currently studying the translation of the mRNA for the enzyme ornithine decarboxylase as an example of insulin-stimulated mRNA translation. She recently found a direct correlation between the ability of insulin to stimulate mRNA translation and the extent of predicted intrachain folding or secondary structure for that mRNA. In searching for a mechanism for this effect, she found that insulin stimulated the phosphorylation—and presumably the activation—of two translation initiation factors whose actions involve unwinding or "melting" those secondary mRNA structures. These studies have led her to propose a general model for insulin-stimulated mRNA translation based on the insulin-stimulated unwinding of mRNA secondary structures.

Insulin also stimulates the transcription of many genes shortly after its addition to sensitive target cells. One such gene, the c-*fos* proto-oncogene, has been studied in detail. The transcription of this gene is stimulated by insulin within approximately 5–10 min of cellular exposure. Recent studies by Rajesh Malik and Michael Roe have identified two cellular proteins that appear to be modified by insulin and growth factor treatment, so that their binding to the regulatory region of the c-*fos* gene is increased. They are currently attempting to purify and clone these proteins, with the goal of determining how insulin stimulates their interaction with the c-*fos* gene promoter. Wi-Sheung Lai has recently cloned another gene whose rapid transcription is stimulated by insulin, the tris-tetraproline (TTP) gene; similar studies of its regulatory elements are under way. The protein encoded by this rapidly turned on gene appears to belong to a novel class of zinc finger–containing proteins that are probably DNA-binding proteins; efforts to determine the DNA targets of this protein are ongoing.

Other protein kinases are involved in the intracellular responses to a variety of polypeptide hormones and growth factors other than insulin. For example, some of these growth factors also

stimulate the rapid breakdown of membrane lipids, generating intracellular signals that activate an intracellular serine/threonine kinase known as protein kinase C, and increase concentrations of intracellular calcium. Activation of protein kinase C leads to the phosphorylation of many intracellular proteins and subsequent metabolic processes, including cell growth, hormone biosynthesis, and neurotransmission. We have been studying one prominent substrate for protein kinase C, the myristoylated alanine-rich C kinase substrate (MARCKS) protein, which is phosphorylated within seconds of protein kinase C activation in intact cells. Recent studies have suggested a role for this protein in regulation of the cellular availability of calmodulin, a calcium-binding protein that can activate a number of important enzymes that respond to changes in intracellular calcium. These studies may lead to the biochemical characterization of a link between the protein kinase C and calcium pathways activated by several polypeptide growth factors.

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. This is a part, transient (or permanent), of the protein to be translocated. The primary structure of numerous representatives for such a sequence has been established in our laboratory and 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, this machinery comprises what has been termed a protein translocon.

Our laboratory is working on translocons for four distinct cellular entities, namely 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 protein from the cytoplasm to the chloroplast interior ("stroma") across outer and inner chloroplast membranes.

From studies on these four translocons so far, but especially on the ER one, it is likely that these and others are 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*. SRF has as its functions *a*) recognition of the signal sequence and *b*) 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 *a*) dissociation of the signal sequence and *b*) its presentation to a signal sequence receptor, which might be a subunit of the protein-conducting channel. This channel would close immediately upon completion of translocation, only to open again after presentation of another signal sequence. The peptidase would, in most cases, remove the signal sequence either during translocation or shortly thereafter.

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 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, 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 pore complex and have recently established the primary structure of a large membrane protein that is specifically located in the pore membrane domain of the nuclear envelope membrane. We are also 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 region 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 nuclear pore. Moreover, we have recently identified a lamin B receptor in the inner nuclear membrane. Its primary structure has been determined by molecular cloning and cDNA sequencing. It 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 the Weinstock Professor of Microbiology and Immunology at the 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 Board. He is a member of the National Academy of Sciences and the Institute of Medicine.

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 unique 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 10–12 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 bacterial pathogen of humans ever described, it remains one of the very few that has never been cultivated in the test tube. Yet its antigens can be produced and studied vicariously in *Escherichia coli* by means of recombinant DNA technology.

Tuberculosis is the major infectious cause of death 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. AIDS (acquired immune deficiency syndrome) causes a breakdown of resistance to tuberculosis and has produced a grave increase in the disease, both in the developing and the industrialized countries.

Immunologic Unresponsiveness and Leprosy

One of the fundamental issues 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 failure 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. It is becoming increasingly clear, however, that not all such clones can be deleted in the thymus and that 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 in humans. 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. Since infection occurs after birth, there is little evidence of clonal deletion of T cells capable of reacting to the 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, in that lepromatous leprosy patients unable to respond to antigens of *M. leprae* usually respond to those of *M. tuberculosis*, which is a very 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 very controversial subject, but the idea that some T cells can down-regulate immune responses, particularly self-destructive ones, is compelling.

Our studies *in vitro* have shown that about 85 percent of responsive 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, we found the suppressor cells to have a unique pattern of antigen recognition different from other cytotoxic or lymphokine-producing T cells. They carry a particular surface marker, CD8, and recognize foreign antigens in association with a particular region, HLA-DQ, 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. We hope to understand the molecular basis of T cell suppression.

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 current vaccines: 1) they require multiple booster shots to be effective; 2) they cannot be given for 6–12 months after birth, because of transfer of maternal antibodies that inactivate them; and 3) the cost. BCG (bacille Calmette-Guérin) is the most widely used vaccine in the world. It is a live, attenuated bovine tubercle bacillus given to protect children against tuberculosis. Studies by our collaborator, Jacinto Convit, in Venezuela, revealed that immunologically unresponsive patients with lepromatous leprosy could be converted to positive cell-mediated immunity and often self-cure by injection of a mixture of killed *M. leprae* together with BCG vaccine. This represents the first instance of vaccines being used therapeutically. It is also the first example of patients overcoming a specific immunologic tolerance, and indicates that BCG is a potent human immune enhancer, or "adjuvant."

BCG has other unique attributes as well. It has been given to over 2.5 billion people and has a very low incidence of serious side effects. It is one of the few 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 six cents 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 undertook to develop new genetic systems for introducing foreign genes into mycobacteria, particularly BCG strains. We developed a shuttle strategy in which mycobacterial DNA could be genetically cloned and manipulated in *E. coli*, 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*, *E. coli*, schistosomes, malaria, leishmania, and HIV (human immunodeficiency virus). 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 in 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 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.

THE long-term goal of our research is to understand how proteins of the extracellular matrix contribute to tissue 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 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 last 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, and the triple helix formed *in vitro* is stable enough that its structure can be characterized by nuclear magnetic resonance (NMR). Therefore the effect of a given mutation can be quantified by directly comparing the behavior of a normal peptide with that of a mutant one. Also, cellular transfection methods have been developed to express and assemble collagen molecules *in vitro*; and 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. This suggests a signal transduction pathway for bone in which the mechanical environment (e.g., strain) influences the pattern of osteoblast gene expression. Particularly intriguing was our observation that the adaptation was associated with a significant improvement in bone strength.

Experiments utilizing other transgenic mouse

strains have been designed to test the hypothesis that adaptations of this type represent a general strategy to treat medical conditions characterized by abnormal bone fragility.

Lymphocyte Differentiation and Activation

Alfred L. M. Bothwell, Ph.D.—Associate Investigator

Dr. Bothwell is also Associate Professor of Immunobiology at Yale University Medical School. He received his A.B. degree in biology from Washington University and a Ph.D. degree in biology from Yale University, where he studied with Sid Altman. This was followed by postdoctoral work at the Cold Spring Harbor Laboratory and at the Massachusetts Institute of Technology with David Baltimore.

THE cellular events that accompany lymphoid differentiation are highly regulated. This regulation of lymphoid gene expression is central to the development of T and B cells and the ultimate function of mature lymphoid cells. We are investigating several experimental systems:

- The involvement of Ly-6 antigens in the development of T cell responsiveness, with a focus on such events as recognition of foreign antigen and processes associated with signal transduction.
- The potential effects of defective Ly-6 gene expression on development of autoimmunity.
- T cell activation, which occurs as a result of specific recognition by the antigen-specific receptor. We have developed an expression system that results in production of high levels of soluble T cell receptors (TCRs) for analysis of function and biophysical properties.
- Regulation of lymphoid-specific genes by RNA splicing. A factor that may regulate this process has been molecularly cloned and is being studied functionally.
- Formation and development of memory B cells and plasma cell lineages. This is studied with an emphasis on commitment to a lineage and its development.

The Ly-6 antigens were initially characterized as associated with activated T cells. The Ly-6A antigen is expressed on these cells and hematopoietic stem cells as well. The antigen can also be used to induce activation, as monoclonal antibodies directed against Ly-6A activate T cells when crosslinked. Activation requires the presence of the TCR complex on the cell surface.

To analyze this process functionally, we are using site-directed mutagenesis. The protein is anchored in the membrane not by a conventional transmembrane segment but by a glycosyl phosphatidylinositol (GPI) lipid anchor. Mutants have been made with an altered anchor, and a transmembrane form has been created. These constructs have shown that the GPI anchor is required for function.

Another form of mutation is the functional deletion of the gene product. When the expression of Ly-6A is inhibited in a T cell clone by antisense RNA, the antigen recognition function of the cell is destroyed and survival of the clone in culture is dependent on certain lymphokines. These experiments reveal the importance of Ly-6 antigens to T cell function and have opened up new approaches to the study of Ly-6 function. We have begun experiments to disrupt the endogenous gene in embryonic stem cells by homologous recombination. These cells can be used to generate a mouse that can transmit the mutation in the germline.

All inbred strains of mice tested display a common pattern of Ly-6C expression except NOD, NZB, and ST. In these three strains, expression of Ly-6C is greatly reduced in the bone marrow and cannot be detected in the spleen or lymph nodes. NOD and NZB mice are well-studied models of autoimmune disease (insulin-dependent diabetes in the former case, systemic lupus erythematosus in the latter). These two strains of mice also exhibit a depressed syngeneic mixed-lymphocyte reaction (SMLR), which, as a measure of the proliferative response of T cells to self-MHC class II determinants, represents an immunoregulatory cellular interaction. Analysis of recombinant inbred lines derived from NZB indicates cosegregation of the defect in Ly-6C expression and decreased responsiveness in the SMLR.

Cloning and sequencing of the affected Ly-6C gene segment from the NOD mouse indicates the presence of an interruption by sequences characteristic of retroposon insertion in the 5′ flanking region of the Ly-6C gene. This disruption of the normal Ly-6C gene allows only a minute level of Ly-6C to be expressed, raising the possibility that the antigen is directly involved in regulatory T cell responses. Ly-6C might be necessary for some aspects of the normal development of the TCR repertoire. We have identified defects in hematopoietic stem cell populations and, using a panel of antibodies, are attempting to identify and purify the cells that can transfer disease.

The Ly-6C and Ly-6A genes are very inducible with interferons. Regions in the promoters of both that confer responsiveness to interferon have been identified but lack known consensus

interferon-responsive elements. We are examining deletion and point mutation analysis of these promoters, using transient and stable transfectants. The segments of DNA that control differences in tissue-specific expression are being defined in transfected cell lines, and the DNA-binding proteins that control the expression of these genes are also under study.

One of the most fundamental interactions in the immune system is that of the TCR with MHC plus ligand. We have produced large amounts of soluble MHC class I– and class II–restricted TCRs to study the molecular interactions. We have been able to inhibit the functional activity of T cell clones by the addition of TCR molecules to block the interaction of T cells and antigen-presenting cells.

Regulation of tissue-specific expression of exons is a major form of gene regulation. We have molecularly cloned the protein that binds the polypyrimidine regions at 3′ splice junctions. This protein may be one of the more significant factors that determine the splicing pattern of a given gene. The region in the protein that contacts RNA contains a novel RNA recognition motif (RRM). We are characterizing this RRM, as well as modifying the expression of the gene product in T cell clones, to determine whether this protein in a cell can alter the splicing pattern of genes.

A primary response generates antibodies and immunologic memory. We are studying the molecular mechanisms that generate B cell memory and the relationship of somatic mutation to that process. Studies of a primary immune response have indicated several new aspects of early events. The early response has an oligoclonal structure, and hypermutation begins earlier than previously recognized. The properties of newly acquired somatic mutations indicate the existence of a differentiation pathway in which mutation is inactive. Once the memory lineages are generated, they continue to mutate. The secondary immunization triggers a new round of proliferation and mutation.

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 the 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 cells. 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 is 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.

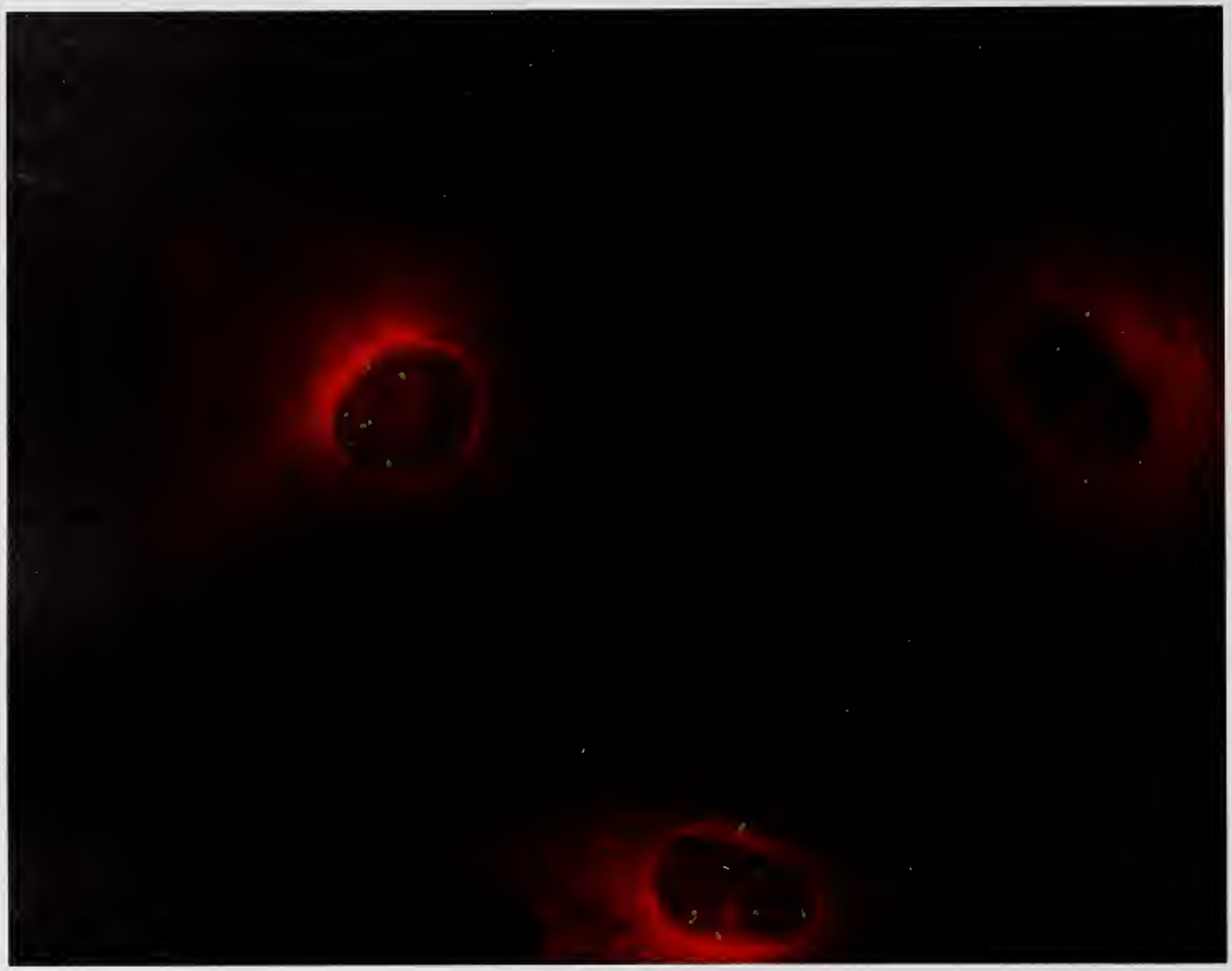
Recent studies in this laboratory have focused on determining whether functionally distinct subsets of CD4 T cells exist *in vivo* and whether the selective activation of one subset or the other has the expected functional consequences. Numerous clinical studies have suggested that various immunization 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, an 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 activa-

tion do subsets of CD4-bearing T cells arise? A particularly interesting question was whether the commitment of CD4-bearing T cells to the Th1 and Th2 subsets was 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.



Tracking the intracellular location of retroviral DNA. To determine how intermediates in the replication of retroviruses move to the nucleus, we can use hybridization probes that specifically recognize the viral DNA and provide a fluorescent signal. This photomicrograph shows cultured rat fibroblasts that have each been infected with about 10 copies of the Moloney murine leukemia virus. The green spots indicate the locations of viral DNA molecules. The nucleus is visualized by staining with a blue fluorescent dye that binds nonspecifically to DNA. An antibody to microtubules, labeled with a red fluorescent dye, indicates the contours of the cytoplasm. By carrying out this kind of analysis at intervals after viral infection, one can track the movement of viral DNA to the cell's nucleus.

Research of Patrick O. Brown.

Retroviral Replication and Human Gene Mapping

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 J. 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 infection can lead to AIDS (acquired immune deficiency syndrome), leukemia, lymphoma, or degenerative diseases of the central nervous system. Millions of people throughout the world are infected with the human immunodeficiency virus, HIV, and will likely succumb to AIDS unless an effective treatment is developed.

Retroviral Replication

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 of the host cell. Thus the viral genome, then called a provirus, becomes an integral part of the cell's DNA. Integration of a provirus into its host cell's genome is essential for retroviral reproduction. This distinctive feature of the life cycle of the retroviruses accounts for many of the characteristics associated with retroviral infection, including insertional mutagenesis, induction of tumors, and viral latency and persistence. Moreover, the fact that retroviruses are designed to introduce foreign genes into cellular DNA makes them exceptionally useful as tools for genetic engineering.

The question that propels much of the work in our laboratory is, How does a retrovirus get its DNA into a cell's nucleus and integrate it into the cell's DNA? To address this question, we are using genetic, biochemical, and cytological approaches. For example, fluorescent probes that bind to viral DNA enable us to track individual viral particles as they infect a cell. This allows us to investigate the influence of such factors as the cell division cycle and the role of specific cytoskeletal elements on the entry of viral particles into the nucleus of cells. The methods we have developed to study such nuclear entry will also be applied in exploring the mechanisms of subcellular localization and the intracellular trafficking of intermediates in viral replication.

Disassembly of the nuclear envelope at mitosis provides one possible route for nuclear entry of viral intermediates. Hence we are studying the role of mitosis in the viral life cycle. 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, but the basis for this phenomenon remains unknown. To bring this phenomenon into clearer focus, we are investigating the dependence of specific steps in the life cycle of the murine leukemia virus (MLV) on the host cell's stage in its own division cycle. Understanding how cellular functions can determine the fate of an infecting retrovirus may lead us to new approaches to antiviral therapy and to improvements in the use of retroviruses as vectors for gene therapy.

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 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. This complex was the focus of our initial studies of integration, and we are continuing to investigate its structure and composition.

Our ability, however, to study the enzymology of integrase, the viral protein that actually catalyzes integration, was limited by the paucity of integrative complexes in infected cells. In the past year, we have constructed bacterial strains genetically engineered to produce abundant quantities of the integrase proteins from HIV and MLV and have developed simple purifications of these proteins. Using model substrates, we can

now readily study the catalytic activities, which include the sequence-specific processing of the viral DNA ends and the joining of these ends to a target DNA molecule.

Studies are under way to define the structure of the integrase proteins, the molecular mechanism by which they carry out their catalytic functions, and the means by which they recognize and bind their two distinct classes of DNA substrates. The ease with which genetic manipulations of integrase can be accomplished in the bacterial expression system will enable us to carry out a genetic dissection of the many activities of the HIV and MLV integrases. Moreover, this system should 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, form 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.

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 the genes by conventional linkage mapping. The same genes could in principle be mapped by an alternative strategy that involves collecting and analyzing pairs of relatives that share a trait of interest. In the case of rare recessive traits, just a few affected inbred individuals could suffice, and such small sets of affected relatives are generally easier to collect than large pedigrees. However, linkage mapping with small sets of relatives generally requires analysis of a large number of closely spaced, highly polymorphic genetic markers, which makes this strategy impractical with current technology.

We are developing a new set of genetic tools that should facilitate widespread application of these highly efficient linkage mapping methods. Experiments are in progress to test our new mapping procedure, using as test systems the nematode *Caenorhabditis elegans* and some well-characterized human families. We hope to be able soon to apply this technology to map genes for a number of complex human traits that have resisted conventional approaches.

Normal and Oncogenic Forms of the src *Gene Product*

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.

IN 1911, Peyton Rous at the Rockefeller University first demonstrated that tumors could be induced by a transmissible agent isolated from a chicken sarcoma. This agent, later shown to be a retrovirus, was designated Rous sarcoma virus (RSV). Since then, RSV has served as a model system to examine tumorigenesis. One of the four genes that RSV carries is both necessary and sufficient to cause tumor formation in chickens. This oncogene, v-*src*, is derived from a normal cellular gene (c-*src*) that was captured by RSV during infection of chicken cells. The c-*src* gene is present in normal chromosomal DNA from humans and from lower eukaryotic species, such as fruit flies and sponges. The high degree of conservation of this gene throughout evolution suggests that it plays a crucial role in normal cell processes. The v-*src* gene from RSV contains many mutations that distinguish it from c-*src*. The *src* gene codes for a 60-kDa protein that is located on the cytoplasmic side of the plasma membrane. The *src* protein, an enzyme referred to as a protein kinase, catalyzes the transfer of a phosphate group from ATP to tyrosine residues on other cellular proteins. The addition of phosphate to proteins is commonly employed by cells to regulate the activity of cellular proteins. It is believed that the mutated RSV *src* protein induces tumor formation by the inappropriate phosphorylation of cellular proteins that are involved in regulating cellular growth. Our major objectives are 1) to determine the function of the *src* protein in normal cells, 2) to understand how this protein is regulated in normal cells, 3) to determine how mutations in the structure of the *src* protein activate the ability of this protein to convert a normal cell to a tumor cell, and 4) to elucidate the mechanism whereby these mutated forms induce tumor formation.

We have found that two types of normal cells express high levels of the *src* protein: 1) neurons from embryonic and adult neural tissues and 2) peripheral blood platelets. To investigate the function of the *src* protein in neurons, we have microinjected antibodies directed against the *src* protein into PC12 cells, a neural tumor cell line that differentiates into neuron-like cells after treatment with nerve growth factor (NGF). These studies were performed in collaboration with Simon Halegoua (State University of New York at Stony Brook).

Introduction of purified monoclonal antibodies to the *src* protein inhibited the production of neurite extensions, the characteristic morphological marker of neuron-like differentiation. Using this antibody, we have begun to examine how *src* communicates with other regulatory proteins believed to participate in NGF-induced neurite outgrowth. We have provided evidence that *src* collaborates with another proto-oncogene, *ras*, to provide essential functions that act on a single pathway involved in NGF-induced neurite outgrowth. We have also found that the oncogenic forms of several tyrosine kinases, in addition to v-*src* (as shown by Alema and co-workers), are able to induce neurite outgrowth in PC12 cells. These studies strongly support the possibility that tyrosine phosphorylation is important in the regulation of critical events in neural cell development.

We have characterized a uniquely modified form of the *src* protein that is specifically produced in central nervous system neurons. This variant form of the *src* protein (src^+), which contains six additional amino acids that are not found in the *src* protein expressed in any other cell type, is generated by an alternative processing of the *src*-specific messenger RNA in neurons. The src^+ protein displays an elevated level of kinase activity. This unique modification suggests that *src* plays a specific role in these specialized cells. Using antibodies that exclusively recognize this unique form of the *src* protein, we have localized src^+ to specific subclasses of neurons in different regions of the brain. Further studies are designed to determine the precise intracellular localization of this form of *src* in developing and mature neurons to provide additional clues to the function of this protein.

Platelets are small anucleate cells in peripheral blood that contain many intracellular vesicles whose components are released upon activation

by cellular hormones. Platelets and the released products carry out specific functions in clot formation and wound healing. Platelets are an ideal system to study cellular events involved in the transduction of extracellular signals. The *src* protein represents a large percentage of total cell protein (0.2–0.4 percent); thus this enzyme may regulate important events at the plasma membrane. We have recently found that four protein tyrosine kinases closely related to c-*src* are also expressed in platelets and that thrombin and other activators of platelet functions cause a rapid increase in the phosphorylation of multiple proteins on tyrosine.

We have found that the phosphorylation of several of these proteins requires platelet aggregation and the interaction between the blood cell adhesion protein fibrinogen and its receptor on the platelet surface. Tyrosine phosphorylation of these proteins is not necessary for secretion of the intracellular platelet granules (the other major function of peripheral blood platelets). These studies suggest that tyrosine phosphorylation may be involved in triggering intracellular events that are induced by fibrinogen-mediated platelet aggregation. We are attempting to purify these proteins, to identify their function in platelets, and to determine how the phosphorylation of these proteins on tyrosine residues affects their functional activity.

The expression of high levels of the *src* protein in terminally differentiated cells, such as neurons and platelets, indicates that c-*src* is not exclusively involved in events that directly regulate cellular proliferation. It is difficult to reconcile the tumorigenic activity of the RSV *src* gene product with such a role for the c-*src* protein in terminally differentiated cells. Mutational changes in the v-*src* gene probably deregulate and activate the tyrosine kinase activity of the v-*src* gene product, thus allowing inappropriate phosphorylation of cellular substrates that are involved in growth regulation. The identification of these substrates is one of the most difficult and important problems in this field.

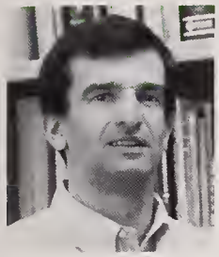
Regulation of the *src* Protein Activation of Tumorigenic Potential

Evidence suggests that the c-*src* protein is tightly regulated in normal cells and that mutational alterations in this protein cause a deregulation and constitutive activation of tyrosine protein kinase activity. We are using site-directed mutagenesis to define regions of the *src* molecule that affect the kinase activity of this enzyme, to identify regions that are important for interaction with other cellular proteins that mediate *src*-induced transformation.

We have focused on the amino-terminal half of the *src* protein that lies outside of the catalytic domain of the molecule, in two highly conserved regions (SH_2 and SH_3) that are shared with several other cellular proteins. Deletion of either of these regions does not prevent transformation by oncogenic forms of *src* but does interfere with the interaction of *src* with several protein substrates. In addition, deletion of the SH_2 or SH_3 domain in wild-type c-*src* activates the kinase activity and transformation potential of this proto-oncogene. We are currently examining the importance of these regions in events other than oncogenic transformation.

We have also identified a novel protein substrate that is phosphorylated in v-*src*-transformed cells. This protein, connexin43, is the major component of gap junctions, which mediate cell-to-cell communication. The phosphorylation of connexin43 correlated with the ability of v-*src* to cause a reduction in junctional communication between cells, suggesting that phosphorylation of connexin43 by the v-*src* protein may be responsible for the inhibition of junctional communication in v-*src*-transformed cells.

Computational Structural Biology



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

Dr. Brünger is also Assistant Professor in the Department 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. 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 set 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. We are also directly combining macromolecular simulation with experimental data to make data analysis possible or more efficient.

Generalizing Molecular Replacement in X-ray Crystallography

In macromolecular crystallography, the determination of initial phases is the major obstacle to determining the structure of the crystallized molecule. This is because the observable x-ray diffraction information from a single crystal comprises only the amplitudes but not the phases of the reflections. Although this "phase problem" has been solved in the case of small molecules (up to a few hundred atoms in the unit cell) through the development of direct methods by Hauptman and Karle, the application of these methods to macromolecules has so far been unsuccessful, and one has to resort to time-consuming and sometimes difficult experimental methods.

The initial determination of phases by molecular replacement is often attempted if the structure of a similar or homologous protein is known. Molecular replacement involves the placement (i.e., rotation and translation) of the known structure in the unit cell of the target crystal in order to obtain the best agreement between calculated model diffraction data and the observed diffraction data. Recent progress in obtaining approximate three-dimensional models of macromolecules from information other than x-ray crystallography suggests an increased use of molecular replacement. For instance, the database of known protein sequences is growing rapidly. Techniques for aligning sequences, such as consensus templates, have been developed to recognize distantly related proteins or protein domains and to carry out model building on the basis of the several hundred known protein structures.

Molecular replacement often fails if the search model is too inaccurate—that is, if the differences in atomic positions between the search model and the crystal structure are more than 1 Å. In this case we proposed to vary the orientations and positions of domains, or structural subunits, of the search model in a neighborhood around the initial positions. We developed a new molecular replacement strategy. First a conventional rotation search is carried out. Then the rotation search is "filtered" by employing refinement of the domain orientations and positions against the Patterson correlation coefficient. Finally, the refined search models with the highest correlation coefficients are used for conventional translation searches. Computer model studies have already suggested the usefulness of the new method. Recently we obtained phases for six crystal structures that previously could not be solved, including several monoclonal antibody Fab fragments.

Presently, the molecular replacement strategy is extended by applying modern techniques of nonlinear optimization, such as simulated annealing.

Predictions of Helix-Helix Association and Stability in Proteins: Leucine Zipper and Membrane Proteins

Prediction of the three-dimensional structure of proteins based on their sequence has up to now been impossible. This fundamental problem of structural biology (the "folding problem") remains unsolved, despite improvements in computational techniques for macromolecular simu-

lation 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 are believed to form coiled coil conformations. Conformational search strategies are being employed with empirical energy functions, using molecular dynamics and energy minimization starting with generic coiled coil–forming α-helices.

Presently we are applying 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 predicted the structure of the dimerization domain of GCN4, for which a high-resolution x-ray structure will be available shortly (personal communication, Thomas Alber, University of Utah). Furthermore, we are trying to explain why Jun can form a homodimeric protein that binds to DNA, whereas Fos is unable to do so and only occurs in a Jun-Fos heterodimer. We are also applying this approach to glycophorin A and other membrane proteins that are believed to form predominantly helical structural elements crossing the membrane. Interesting results were obtained during the past year for vacuum simulations. Currently we are attempting to incorporate the environment (water and lipid molecules) into the simulation. A rapidly increasing amount of experimental data on mutants of the systems studied will allow a thorough comparison of the theoretical predictions and the experimental results.

Macromolecular Simulation of Free-Energy Differences

We are involved in a number of projects that are aimed at simulating free-energy differences between two states of a biological system, using the free-energy perturbation technique. Our goals are to investigate microscopically the structure and stability of protein secondary structural elements and protein-peptide complexes and to evaluate the reliability of free-energy calculations and molecular dynamics simulations as tools for this purpose. In particular, we are studying 1) 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, Julian Sturtevant, and their colleagues (Yale University) and 2) the complexes of L-tryptophan and several of its chemical analogues with *Escherichia coli trp* aporepressor, for which x-ray crystal structures and binding free energies have been measured by Paul Sigler (HHMI, Yale University) and his colleagues.

We also plan to carry out free-energy calculations on mutants of β-turns in staphylococcal nuclease, in a joint project with Robert Fox (HHMI, Yale University). High-resolution x-ray structures of various mutants and thermodynamic data through temperature-dependent NMR studies have been obtained. Specifically, we would like to simulate the free-energy differences between the cis and trans conformations of Pro-117 for a number of point mutations in the vicinity of the proline residue.

Biophysical Studies of Eukaryotic Gene Regulation and Molecular Recognition

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

Dr. Burley is also Assistant Professor 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. The systems we have chosen to study are models for gene regulation in eukaryotes and slow-binding inhibition of enzymes.

Our approach is to use the methods of x-ray crystallography to determine the three-dimensional structures of biological macromolecules and their complexes with DNA or other ligands. These structures contain a wealth of atomic detail that can be analyzed 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 structural studies and analyses will allow us to exploit the powerful formalism of physics to classify systematically the interactions between individual atoms that effect 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 have begun to examine two model systems of eukaryotic gene regulation, with the goal of improving our understanding of the structural and physical bases of transcriptional control of genes. First, we are collaborating with Robert Roeder (the Rockefeller University) on x-ray crystallographic studies of transcription factor IID (TFIID) and upstream stimulatory factor (USF). These two proteins are involved in transcription of class II genes in eukaryotes. TFIID binds to the TATA consensus sequence and functions as a general transcription initiation factor. USF is a member of the c-*myc*-related family of DNA-binding proteins and contains both a helix-loop-helix motif and a leucine repeat. It binds as a dimer to an upstream activating sequence GGCCACGTGACC. During transcription, TFIID and USF bind to DNA in close proximity and interact with each other to enhance both DNA binding and transcription. In addition to determining the three-dimensional structures of these proteins complexed to their respective promoter elements, we hope to determine the structure of a complex consisting of TFIID, USF, and DNA.

Second, 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, which bears no similarity to previously defined DNA-binding motifs and is thought to represent an entirely new type of DNA-binding protein.

Successful three-dimensional structure determinations of these model systems will provide insights into three different modes of DNA-protein interaction and may also give some information about intermolecular interactions between proteins bound to the same piece of DNA.

Slow-binding Enzyme Inhibition

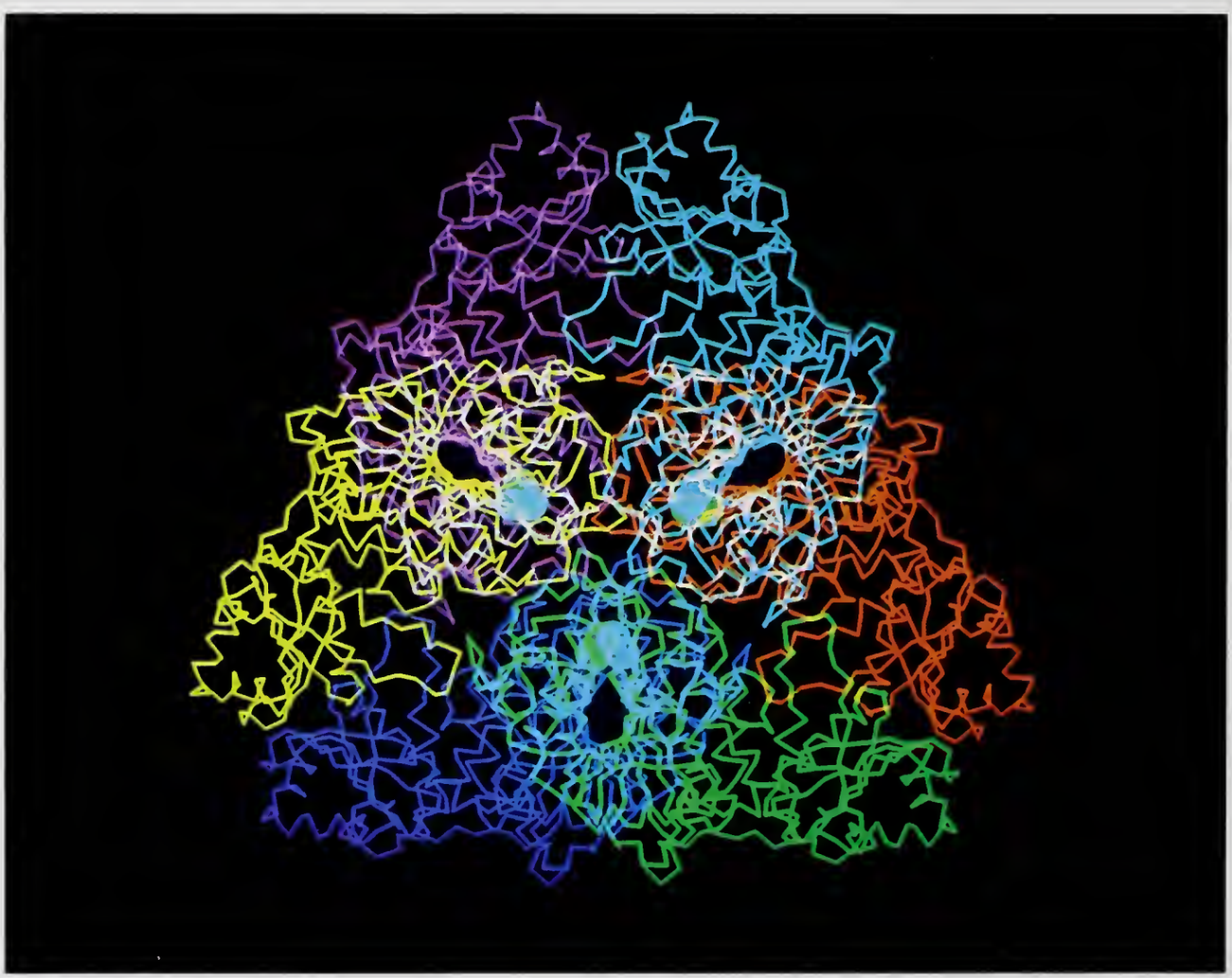
My previous work in William Lipscomb's laboratory (Harvard University) employed x-ray crystallography to determine the first three-dimensional structure of an amino-terminal exoprotease, a two-zinc metalloenzyme known as leucine aminopeptidase (LAP) isolated from bovine lens. We discovered that the enzyme consists of two unequal α/β domains. In its active form it exists as a hexamer, which resembles the allosteric enzyme aspartate transcarbamoylase in appearance. The enzyme's active site consists of a novel bimetallic cluster, with the two zinc ions separated by 2.9 Å and coordinated only by carboxylate and carbonyl oxygen atoms. The precise structure of the active site allowed us to suggest a mechanism by which the enzyme may effect peptide bond hydrolysis.

In addition, we determined the isomorphous

structure of the enzyme complexed with its slow-binding inhibitor bestatin, which is a dipeptide analogue that lacks a scissile peptide bond. A detailed analysis of the interactions between bestatin and the active-site residues allowed us to suggest a mechanism of slow-binding inhibition of LAP and other bimetallic, amino-terminal exoproteases.

Finally, the three-dimensional structure of LAP also provides some insights into the role that post-translational modification plays in the problem of recombination. LAP is homologous to *pepA*, a manganese-dependent, hexameric aminopeptidase found in *Escherichia coli* that is required for plasmid ColE1 site-specific recombination.

We will continue similar work with other enzyme systems of biological and biochemical importance.



A computer graphics representation of the three-dimensional structure of the hexameric, zinc-dependent leucine aminopeptidase from bovine lens that has been determined to 2.25 Å resolution. This enzyme is responsible for post-translational modification of proteins and oligopeptides and for degradation of aging proteins in a diverse group of organisms ranging from Escherichia coli *to humans.*

Research of Stephen K. Burley.

Molecular Studies of Ca^{2+} 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. 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.

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

Over the last nine 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.

Excitation-Contraction Coupling in Skeletal Muscle

Muscle contraction is initiated by a depolarization of the transverse tubular membrane, which in turn signals the release of Ca^{2+} from the junctional sarcoplasmic reticulum. One specific aim of our research is to understand the structure and function of protein components of the sarcoplasmic reticulum membrane. We have purified the ryanodine receptor of rabbit muscle sarcoplasmic reticulum and have shown that it can mediate single-channel activity identical to that of the sarcoplasmic reticulum's Ca^{2+} release channels. The morphology of the purified receptor has revealed that it is identical to the "SR feet" and thus would seem to play a dual role in excitation-contraction coupling as the Ca^{2+} release channel and the bridging structure in the junctional gap.

A second specific aim of this research concerns the dihydropyridine-sensitive Ca^{2+} channel of skeletal muscle and its dual role as a voltage sensor for excitation-contraction coupling and a Ca^{2+} channel. The dihydropyridine receptor has been purified from rabbit skeletal muscle and shown to consist of four subunits—α_1, α_2, β, and γ. In the last year we determined the structure of the γ-subunit by isolating cDNAs from an expression library. We deduced that the primary structure is a 25,058-dalton protein containing four transmembrane domains and two N-linked glycosylation sites. This description is consistent with biochemical analyses showing that the γ-subunit is a glycosylated hydrophobic protein.

We have also analyzed the dihydropyridine receptor's α_2-subunit and associated δ peptides. The results have shown that the δ peptides are not true subunits, but instead represent the carboxyl-terminal peptide of a proteolytically processed α_2-subunit. The goal of this project now is to determine how the receptors for dihydropyridine and ryanodine are coupled in the excitation-contraction process.

Neuronal Ca^{2+} Release Channels

A second area of research in our laboratory concerns Ca^{2+} release channels in neuronal cells. Inositol 1,4,5-trisphosphate (IP_3), produced after receptor activation, is an important second messenger in the release of stored Ca^{2+} from intracellular compartments. Although IP_3 appears to have a major role in Ca^{2+} regulation within neurons, physiological and pharmacological evidence has indicated the presence of non-IP_3-gated Ca^{2+} pools. We have identified a high-affinity ryanodine receptor in rabbit brain membranes as a 400-kDa protein that cross-reacts with antibodies against the sarcoplasmic reticulum ryanodine receptor and migrates as a large oligomeric complex of about 30S on sucrose density gradients. Although we have not identified the exact physiological role of the brain ryanodine receptor, we are now attempting to purify it so that it may be shown responsible for the non-IP_3-gated Ca^{2+} release in neurons.

Molecular Pathogenesis of Muscular Dystrophy

A third area of research concerns the structure and function of dystrophin, the high-molecular-weight protein product of the human Duchenne muscular dystrophy (DMD) gene. Dystrophin occurs in the sarcolemma (fiber membrane) of nor-

mal skeletal muscle, but is absent from the muscle of DMD patients and *mdx* mice. The predicted amino acid sequence of dystrophin suggests that it is a membrane cytoskeletal protein and, furthermore, that DMD is a disease of the membrane cytoskeleton.

Studies of other genetic diseases involving the cytoskeleton have shown that a deficiency in a major cytoskeletal component can severely affect the integrity of the cell membrane. For example, spectrin deficiency results in fragile erythrocyte membranes and thus severe hemolytic anemia.

In comparison, dystrophin has not been considered to play a major structural role in the membrane cytoskeleton of skeletal muscle because of its low abundance of approximately 0.002 percent of total muscle protein. In the past year we have determined the cytoskeletal properties and relative abundance of dystrophin in a highly purified preparation of skeletal muscle sarcolemma. We have found that it constituted fully 2 percent of the sarcolemma protein. This suggests that it does play a major structural role and that the integrity of DMD muscle sarcolemma, through lack of dystrophin, is likely to be compromised.

In order to understand the molecular pathogenesis of DMD, it is imperative to identify the proteins that are associated and/or linked to dystrophin, since the absence of one component of the membrane cytoskeleton is sometimes accompanied by the loss of another. In the past year our laboratory has succeeded in purifying a dystrophin-glycoprotein complex from rabbit skeletal muscle and identifying five dystrophin-associated proteins, including four glycoproteins.

In addition to dystrophin, the complex contains a 59-kDa protein triplet and four glycoproteins of 156, 50, 43, and 35 kDa. We have begun to characterize the status of these proteins in dystrophin-free muscle. One interesting finding was a dramatic (about 90 percent) deficiency of the 156-kDa dystrophin-associated glycoprotein in muscle from *mdx* mice and DMD patients. Thus the marked reduction of the 156-kDa glycoprotein in dystrophic muscle, and possibly of other dystrophin-associated proteins, may be the initial step(s) involved in the molecular pathogenesis of muscular dystrophy.

Our goal for the next year is to clone each of the dystrophin-associated glycoproteins and determine their structure and possible function. The results should help to define the function of dystrophin and explain how its absence results in DMD.

Gene Targeting

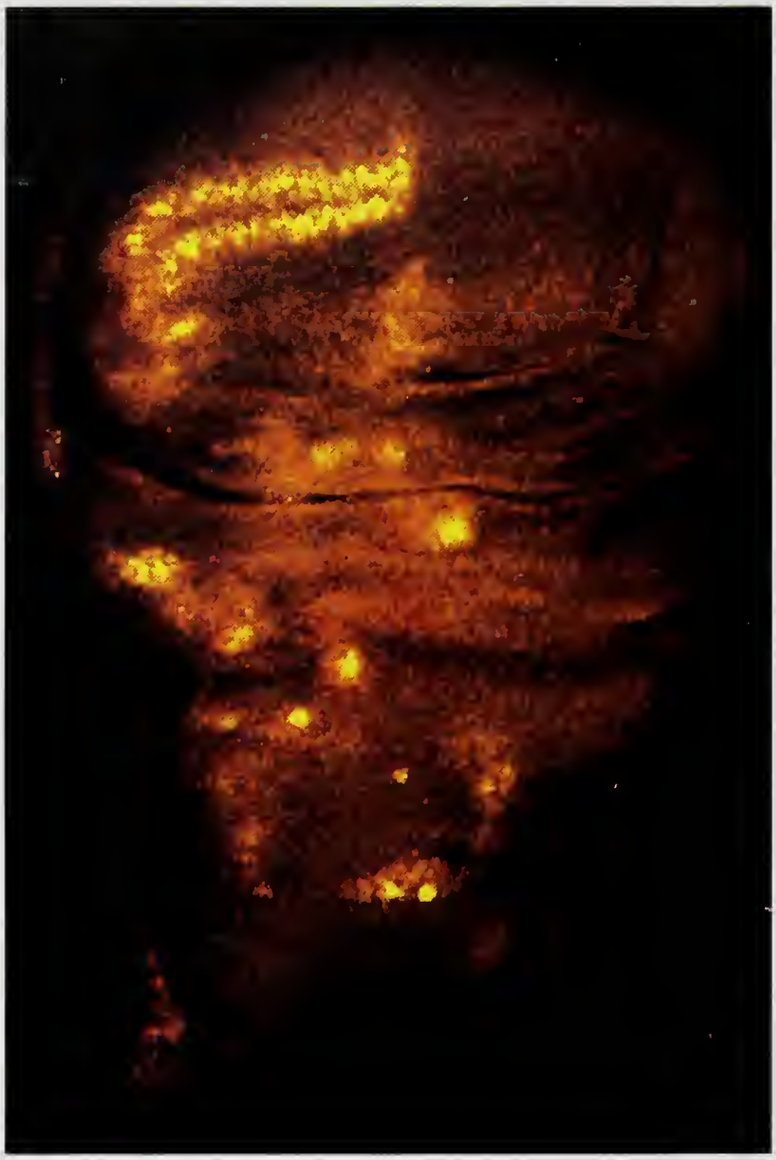
Mario R. Capecchi, Ph.D.—Investigator

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 was recently elected to the National Academy of Sciences.

HOMOLOGOUS recombination between DNA sequences residing in the chromosome and newly introduced DNA sequences, termed gene targeting, provides the means for specifically modifying any gene in any desired manner in cultured mammalian cells. The desired alteration is first introduced into a chosen cloned DNA sequence, using standard recombinant DNA technology. Gene targeting is then used to transfer the gene modification to the cognate gene in the genome of cultured cells. If the recipient cell for the gene modification is a pluripotent, mouse embryo-derived stem (ES) cell, then the means are available for creating chimeric mice from these cells that will transmit the altered gene to their progeny. With this technology the biological function of any cloned gene can be determined.

Our laboratory is using this technology to determine the function of two sets of genes in the early development of the mouse. The first set of genes includes members of the *int* proto-oncogene family (*int-1*, *int-2* . . .), which are believed to be involved in localized developmental decisions through cell-cell signaling. This family of genes contains at least 20 members. Targeted disruption of the murine *int-1* proto-oncogene resulted in mice with severe abnormalities in midbrain and cerebellar development, indicating a prominent role for the *int-1* protein in the induction, from the neural tube, of the mesencephalon and metencephalon. Some of these mice survive birth and exhibit severe ataxia (i.e., loss of balance and coordinated movement). Surprisingly, despite missing a large portion of their midbrain, these mice respond fairly normally to stimuli of light, sound, smell, and touch. We have also created mice with null mutations in *int-2*. Mice heterozygous for this mutation appear normal and are fertile. We are in the process of analyzing the phenotype of *int-2* homozygotes.

The second set of genes that we are analyzing are transcription factors participating in a developmental program for specifying positional information along the anteroposterior axis of the early mouse embryo. These mammalian genes are homologues of the *Drosophila Ultrabithorax* and *Antennapedia* genes. In invertebrates these genes specify the identity of cells within each parasegment. The function of the corresponding genes in the human and mouse is not known. However, their expression pattern argues for an equally important role in mammalian development. In the human and mouse this set of genes, known collectively as the *Hox* genes, contains at least 30 members. We have used gene targeting to disrupt specifically a dozen of these genes in mouse ES cells. These cells are in turn being used to generate chimeric mice that transmit the mutant gene to their progeny. Disruption of the *hox-1.5* gene resulted in mice with multiple, but regionally restricted, developmental defects: *hox-1.5*$^{-}$*/hox-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 muscle tissue. In addition, these mice often feature defects of the heart and arteries, as well as craniofacial abnormalities. This collection of deficiencies is remarkably similar to the pathology afflicting humans with DiGeorge syndrome, a congenital disorder.



Prepattern of sensory organs in imaginal wing disc of Drosophila. *The bright spots are cells expressing the* achaete *protein, which will give rise to adult sensory structures.*

From Skeath, J.B., and Carroll, S.B. 1991. Genes Dev *5:984–995.*

Genetic Control of Segmentation and Segmental Pattern Formation in Drosophila

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

Dr. Carroll is also Assistant 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 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 that are now under evaluation as potential pharmaceuticals. His honors include the NSF Presidential Young Investigator Award.*

VIRTUALLY every animal species can be distinguished by its external and/or internal appearance, its morphology. Gross similarities in external or internal organization usually reflect an evolutionary relationship between animals. Widespread similarities in body organization may reflect a common evolutionary origin of a large number of well-diversified species. For example, a segmentally organized body plan, present in annelids (e.g., earthworms), arthropods (lobsters, spiders, insects), and to a lesser degree in vertebrates (snakes, mice, humans), is one of the most general forms of organization in the animal kingdom. Evolutionary biologists believe that modifications of the basic repeating segmental pattern found in earthworms and millipedes, for example, have led to the diverse array (perhaps more than 1 million species) of arthropods found today. For instance, it is thought that a gradual reduction in leg number on the more posterior segments of a many-legged animal has led to the present-day typical insect body plan, with only three pairs of legs protruding from the thorax followed by a legless abdomen.

To understand how body patterns evolve, we must first learn about how body plans develop. To this end, we and several other HHMI laboratories are engaged in a detailed study of the genetic program that guides the development of the fruit fly *Drosophila melanogaster*. We focus on those genes that are critical to the organization of the segmental outline of embryo and adult, controlling the overall pattern and function of body segments. The wealth of genetic information gathered over the past 70 years and the explosion of *Drosophila* embryology and molecular biology over the past 10 years make it the best complex animal model for understanding the genetic regulation of cell behavior and the three-dimensional organization of tissues, organs, and entire organisms.

Gene Activity During *Drosophila* Development

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 taking place in different regions of the animal that foreshadow the cellular events to follow. For example, in the *Drosophila* embryo, certain key proteins come to be expressed in stripes, which represent the segmental divisions to form later. The second phase of segmental pattern formation involves the establishment of a segmentally repeating array of stem cells within the different germ layers that give rise to the different tissues of the animal. For example, genes such as those of the *achaete-scute* complex (AS-C) are activated only in the stem cells of the central and peripheral nervous system. Finally, in the third phase, these stem cells divide, giving rise to the full complement of differentiated cells that make up different tissues and organs and express distinct structural genes to carry out their specialized tasks.

From molecular prepattern to stem cell formation 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, how it operates and is regulated. From the level of overall body organization to the molecular details of gene regulation, we want to know which genes are key to pattern regulation, how they coordinate gene expression and cell behavior, and how they and the regulatory programs they belong to have changed in the course of the morphological diversification of animals.

To this end, our laboratory focuses on three aspects of the genetic control of pattern formation during *Drosophila* development. First, we are working on the molecular regulation of the pair-rule genes, the first genes expressed in a segmentally repeating prepattern. Second, we are studying how the early embryonic prepattern genes governing segment number, segment polarity, segment identity, and germ-layer specificity

regulate the expression of the stem cell–specific proneural genes of the AS-C. Finally, we are working to understand better the prepattern that governs 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.

Regulation of Segmentation Gene Expression

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 per every two segments. Previous studies have shown that the generation of some pair-rule patterns is derived directly from other such patterns, i.e., stripes carving more stripes, but a select group of pair-rule patterns is generated from the aperiodic pattern of the gap proteins, a small number of transcription factors expressed in broad partially overlapping zones of the early embryo. The *hairy* gene is one pair-rule gene regulated in this way. Our laboratory is trying to determine how the many gap genes carve the *hairy* pattern into stripes. One unexpected observation is that several of the gap proteins appear to act by repressing *hairy* transcription in discrete regions of the embryo, as opposed to activating specific stripes. The elucidation of the direct activators and repressors of *hairy* transcription and their sites of action within the upstream regulatory DNA of the *hairy* locus is a major goal of our present work.

Segmental Pattern Regulation and Proneural Gene Expression

The stripes of pair-rule gene products are not sufficient to specify the precise location of individual cells of any tissue or organ. Rather, multiple regulatory systems that determine segment number, polarity, identity, and germ-layer position must all converge upon subsets of other genes to specify the three-dimensional coordinates of individual cells. To achieve a better understanding of this process, we have begun a detailed study of the spatial regulation of two genes of the AS-C, which promote neural development. Once the early prepatterns are in place, the positional information they provide determines where the genes of the AS-C will be active. We have found that the products of the *achaete* and *scute* genes, which themselves are regulatory proteins, are expressed at distinct dorsoventral positions of the neuroectoderm in both a segmentally repeating and segmentally modulated pattern. The identification of the direct regulators of AS-C gene expression and the sites through which they act is a second major thrust of the laboratory.

Prepatterns in Imaginal Discs

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 appendages are derived from these sacs of cells, which undergo extensive changes during larval growth and pupation. Many of the genes that control the pattern of the relatively small numbers of cells of each embryonic segment also control the final morphology of the adult. The molecular prepatterns that govern eye, wing, leg, and antennal development are not well known. Our laboratory has embarked on a detailed study of certain genes expressed in imaginal discs, such as *hairy*, the AS-C, and others, in order to define better the imaginal prepatterns and the morphogenetic processes they control.

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 the NIH under the supervision of Marshall Nirenberg. Dr. Caskey is a past president of the American Society for Human Genetics and was recently named Distinguished Service Professor by the Board of Trustees of Baylor.

MOLECULAR genetics offers unprecedented opportunities for correction of single-gene defects, the development of simple DNA-based diagnostics, and the discovery of disease genes. This laboratory has made significant progress in the development of gene replacement therapies for three diseases. The isolation of the fragile X locus has been achieved as a collaborative effort with two laboratories. DNA-based methods for diagnosis of affected males and carrier females are now fully developed for four diseases inherited through the X chromosome.

Genetic Correction of Inherited Disease

Human genes can now be cloned, placed into defective viruses used as vectors, and transferred into other cultured cells, embryonic cells, and mice. 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 provides a once-per-lifetime cure, but carries the potential complications of graft-versus-host reaction. PEG-ADA administration (ADA attached to polyethylene glycol) on a continuing basis has provided improvement in immunologic function of patients, but not a cure. We reported previously our high-efficiency retroviral delivery of a human ADA minigene to mouse stem cells that, when returned to a lethally irradiated mouse, resulted in 99 percent rescue and high-level, long-term expression of the gene. Thus stem cell infection has been achieved in mice, and recent human experiments are equally encouraging.

Both "safe" virus and helper cell lines were used to deliver the human ADA gene to normal and ADA-deficient long-term bone marrow cultures. The infection rate exceeds 90 percent, and the level of ADA expression is equal to normal myeloblastic progenitors. This success is a consequence of precisely defined conditions of stem cell stimulation and growth. The parameters include cocultivation of human bone marrow cultures with viral producer lines and selected recombinant interleukin cell-stimulatory reagents. Presently the research is focused on the use of CD34 monoclonal selected cells. These have stem cell properties and represent about 10^{-4}–10^{-5} of human bone marrow cells. If transfer can be mediated by these selected cells, the current large cell-culture requirements for the human experiment are reduced to a trivial number of cells. Clinical protocols are proposed for 1991.

Ornithine transcarbamylase (OTC) deficiency, the most common urea cycle defect in humans, is inherited in an X-linked recessive manner. The 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 graft-versus-host reaction. This condition is poorly managed through dietary protein restriction and medical therapy. Two mouse models, *sparse fur* and *ash*, are available for developing gene correction technology. Using a human OTC minigene under regulation of a small bowel liver-specific promoter, the *sparse fur* mouse has been totally corrected of the deficiency, including both coat features and metabolic defects.

Surprisingly, the correcting enzyme was expressed in small bowel (an ectopic tissue) and not in liver. This observation has directed our efforts for human correction to two target organs: liver and small bowel. Toward that objective, two viral vectors are now available and are found to transfer human OTC successfully to cells in culture. For liver delivery, we use a retroviral vector with tissue-specific promoter. For small bowel delivery, we use a defective adenoassociated virus. Each has biosafety features and delivers genes efficiently. Our studies now focus on the correction of the *ash* mouse mutant, which has profound OTC deficiency. All human studies currently involve liver cells in culture.

Duchenne muscular dystrophy is a severe disorder also inherited in an X-linked recessive

manner. Deficiency of the protein dystrophin leads to multiple abnormalities of muscle dysfunction and death. The gene is large and its mRNA complex. Three mouse models of DMD are available for study. We have recently developed methods for the isolation of a full-length mouse dystrophin minigene that functions normally in cells and mice. Using special muscle-specific promoters controlling the dystrophin minigene, correction of the deficiency in the *mdx* mouse model has been achieved.

Current research efforts on DMD feature mouse and human studies. Our direction in the former studies is to develop efficient methods for gene delivery to muscle cells of the adolescent *mdx* mouse. In the human studies we have embarked on a collaborative effort (with Helen Blau, Stanford University) to isolate DMD muscle satellite cells from very young patients. These cells will be the target for gene delivery and will be returned to the patient.

Gene Isolation and Molecular Diagnostics: Fragile X Locus

The end of the long arm of the X chromosome contains some 40 disease genes, including those for two mental retardation disorders (fragile X syndrome and the oculocerebrorenal syndrome of Lowe) and a severe muscular dystrophy (Emery-Dreifuss). Yeast artificial chromosomes (YACs) containing 80 percent of this 40-megabase region have been isolated and analyzed for possible involvement in these disorders. Clones containing the region involved in Lowe's syndrome and fragile X have been obtained, and the identification of the genes involved is under way.

In the case of fragile X, the YAC identified has allowed the development of a new marker, based on the polymerase chain reaction (PCR), that improves fragile X diagnosis; narrows the region responsible for the syndrome to about 200 kb; and provides the material for deciphering the nature of this most common form of inherited mental retardation. Molecular diagnosis of fragile X now exceeds the accuracy of cytogenetic studies and has been used already at Baylor in three prenatal cases.

Gene Isolation and Molecular Diagnostics: Duchenne Muscular Dystrophy

Simultaneous (multiplex) PCR amplification of multiple regions of the DMD gene is now possible with 81 percent accuracy for deletions and duplications, as determined by a cooperative study conducted at 14 international laboratories. Genetic linkage analysis is developed and tested for a new multiplex amplification method, which employs short tandem repeat polymorphisms (STRs) at six positions in the gene. These developments make scanning for new mutations and genetic linkage efficient, cost-effective, and accurate. The techniques apply to both affected males and female carriers.

Gene Isolation and Molecular Diagnostics: Lesch-Nyhan Syndrome

All Lesch-Nyhan syndrome diagnoses for both affected males and female carriers are performed at the molecular level by direct automated sequencing of genetic defects. The mechanisms of mental retardation and choreoathetosis (involuntary movement), associated with Lesch-Nyhan syndrome, are being studied with transgenic technology. These studies will produce mice deficient in both HPRT (hypoxanthine guanine phosphoribosyltransferase) and uricase, as is the case in the human syndrome. Humans and other hominids appear to have lost uricase through peptide chain termination and splice junction mutations that occurred approximately 20 million years ago.

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 the many honors he has received are the Lasker Award and the 1989 Nobel Prize in chemistry.

A cell must orchestrate thousands of chemical reactions in order to live, to grow, and to respond to its environment. These chemical reactions rarely happen spontaneously, but are usually catalyzed by macromolecules called enzymes. It was long thought that all enzymes were proteins. More recently we and others have found that RNA, a form of genetic material, 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 in part catalyzed by RNA. Second, the study of how RNA enzymes work may reveal hitherto unknown mechanisms of biologic catalysis. Third, RNA enzymes (ribozymes) can be used as sequence-specific RNA cleavage agents *in vitro*, providing useful tools for biochemical studies of RNA. Finally, on a more speculative note, RNA catalysis has the potential of providing new therapeutic agents. For example, it has been suggested that ribozymes directed against viral RNA sequences might be able to cleave and thereby 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.

In the past year we have 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 famous DNA double helix), the ribozyme also binds two of the sugar groups that form the "backbone" of the RNA substrate chain. We expect that this type of recognition will be widespread in biology. In a separate study, we used genetic engineering to introduce small changes near the active site of the ribozyme. We were able to improve greatly both the speed with which the ribozyme cleaves RNA and its specificity (its ability to cleave the correct RNA sequence while leaving others untouched).

In the structural area, we developed a method for monitoring the folding of the RNA chain of the ribozyme. This allowed us to understand one of the roles of magnesium ions in RNA catalysis. We are now using the same technique to explore the contributions of individual structural domains and even individual units (nucleotides) to the folding of the ribozyme.

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 the ciliated protozoan *Oxytricha nova*, because it has 26 million miniature chromosomes per cell. This gives us a large amount of telomeric protein to study. Similarities in the DNA sequences of 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 isolated and sequenced the genes encoding the two subunits of the *Oxytricha* telomere-binding protein. We have now used genetic engineering methods to produce large amounts of these protein subunits in bacteria. The protein synthesized and purified from bacteria forms the same telomeric DNA-protein interaction seen in living *Oxytricha*. Thus we now have a laboratory system for studying telomere structure and function that is convenient and appears to be faithful to the biological system.

We have also begun to search for proteins with

a similar function in human cells and have obtained some promising initial results. Telomeres appear to help organize chromosomes within the nucleus, and this organization has been reported to be perturbed in cancer cells. Information about the human telomere-binding protein and its interaction with the nuclear envelope may therefore be of medical importance.

Molecular Genetics of Histocompatibility

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

Dr. Chaplin is also Assistant 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 medical residency in internal medicine at Parkland Memorial Hospital, Dallas, he received postdoctoral training in genetics at Harvard Medical School with Jonathan Seidman.

THE immune system acts in a critically important fashion to protect the organism from a vast array of infectious and parasitic agents. To execute this function the immune system has evolved potent mechanisms to eliminate or destroy invading organisms. Because these effector mechanisms are so potent, stringent regulation of the activation of the immune response is very important. Inappropriate activation of immune effector systems or failure of the immune system to distinguish self tissues from invading pathogens can lead to serious damage to the host. Examples of such failures of normal immune regulation are insulin-dependent diabetes mellitus (in which the immune system participates prominently in the destruction of the insulin-producing cells of the pancreas), rheumatoid arthritis (in which abnormal self-reactive antibodies cause destructive inflammation in joints), and systemic lupus erythematosus (in which a broad spectrum of self-reactive antibodies induce inflammatory damage in many tissues).

Proper regulation of immune defenses depends on the coordinate action of an array of different host mechanisms. Many of these mechanisms are influenced in fundamental ways by genes of the major histocompatibility complex (MHC). Our laboratory studies the structure and molecular biology of this important gene complex. These studies are motivated by a desire both to understand the nature and regulation of the genes in the MHC and to use the MHC as a model system to test methods for analysis of large gene complexes in general.

The MHC encodes genes that determine whether tissue grafts between unrelated individuals are accepted or rejected. The products of these genes determine each individual's tissue type (also called HLA type). These HLA molecules also serve other important, perhaps more biologically relevant, functions, as they participate in the regulation of most immune reactions.

The MHC comprises a cluster of at least 60 closely linked genes, which can be divided into three classes on the basis of their structures and their functions. Class I genes encode proteins that are present on the surfaces of all nucleated cells, where they participate in a number of immune responses. Importantly, when a host cell is infected with a virus, fragments of viral proteins associate with the class I molecules. This complex of viral protein and class I molecule is recognized by host cytotoxic T lymphocytes, which then act to destroy the virally infected cell.

Class II MHC genes encode cell surface proteins that share some structural features with the class I molecules, indicating evolution from a common ancestral gene. Class II molecules are present on the surfaces of only a few cell types, where they also bind fragments of foreign antigens. In this case the complex of foreign antigen and class II molecule is recognized by host helper T lymphocytes, resulting in the initiation of an active immune response to those antigens.

Class III genes encode molecules with diverse, and in some cases unknown, functions. Four of the class III genes specify soluble proteins that circulate in the blood, where they form part of the most primitive portion of the immune system, the complement system. They provide an initial, nonspecific barrier to invading microorganisms, before a specific immune response is developed. In addition to the complement genes, this region contains the genes encoding two potent, white blood cell–derived inflammatory molecules (tumor necrosis factor and lymphotoxin). The class III region also contains several additional genes, some with as yet unknown functions.

An important feature of most of the MHC genes is their marked variability in structure between different individuals in the population. This individual-to-individual (or allelic) variation confers an immunologic fingerprint that makes up each individual's tissue type and is the basis for recognition and rejection of foreign tissue grafts. More generally, this fingerprint permits very sensitive discrimination of self from nonself.

It has been apparent for over a decade that susceptibility to a large number of immunologically mediated diseases is determined by an individual's tissue type. In some cases, disease susceptibility appears to be conferred by an allelic variant of a single MHC gene. In other cases, susceptibil-

ity seems to be conferred by a complete tissue type (made up of the full complement of class I, II, and III genes). In most cases the mechanism by which these inherited MHC genes affect disease susceptibility is not known. To understand the association of certain MHC alleles with disease susceptibility, it will be necessary to analyze the molecular structures of complete MHC complexes containing those alleles. Such a detailed analysis will also be required for an understanding of the mechanisms and forces that govern the evolution of the complex. Because this evolution determines our repertoire of immune responsiveness, understanding these evolutionary forces will be important for understanding how we adapt to a world of constantly changing pathogens, from simple viruses to complex parasites.

Our laboratory is defining the structure of the entire MHC by molecular cloning. Large fragments of human chromosomes are propagated in yeast following fusion to simple yeast chromosomal elements. These recombinant molecules exist in yeast as independent chromosomes (yeast artificial chromosomes [YACs]). The use of these YACs permits us to analyze DNA fragments 5–10 times larger than previously possible with conventional gene-cloning methods. This is particularly important for analysis of the structure of the MHC, because of the large size of the complex. Our analysis has shown YAC cloning to be a powerful methodology. All but 3 percent of the HLA complex has been isolated in 52 YACs. These molecular clones are stable and reproduce faithfully the structure of the original human chromosome from which they derive. Their structure can be analyzed at high resolution. New DNA probes can be rapidly isolated from the YACs to be used for comparison of the structures of normal and disease-associated MHCs. We anticipate that they will facilitate the identification of additional genes within the gene complex. Analysis of such genes may be critical for an understanding of the mechanisms of disease association.

These studies provide access to the entire structure of the MHC. They also provide an example of the power of the new YAC-cloning technology that may facilitate its general application to the mapping of the human genome and to more extensive analysis of other specific human disease loci.

Hormonal Regulation of Gene Expression

William W. Chin, M.D.—Investigator

Dr. Chin is also Associate Professor of Medicine at Harvard Medical School and 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 the NIH and with Joel Habener at the Massachusetts General Hospital. 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 specific tissues in the body, effect extracellular communication, and regulate cellular function. Our studies have focused on the hormonal regulation of gene expression and on several genes that encode the hormones themselves.

Thyroid Hormone Regulation of TSH Gene Expression

Thyrotropin (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, T3 and T4. These are modified amino acids that regulate metabolism and gene expression in almost every cell of the body. To maintain a constant level of T3 and T4 in the bloodstream, these hormones act back on the pituitary to decrease TSH production and secretion. Hence TSH and thyroid hormones are involved in a classic negative-feedback relationship.

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.

Thyroid hormones, in general, act at the cellular level by entering the cell either as T3 or T4. T4 may be converted to T3 in certain tissues, such as the pituitary. T3 then enters the nucleus, where it interacts with the thyroid hormone receptor (TR), a protein encoded by the proto-oncogene c-*erbA*. The T3-TR complex then binds directly to cis-DNA elements within thyroid hormone–responsive genes to activate the appropriate responses.

Using deletion-mutation analyses of the regulatory regions of the subunit genes, we have localized putative thyroid hormone response elements (TREs) to the 3′ end of the first exon and to a region just 5′ to this exon in the rat TSHβ gene, and to a proximal upstream promoter region of the rat α gene. There is limited sequence similarity between these cis-DNA elements and a TRE defined in the rat growth hormone gene. Consistent with their roles as TREs, these DNAs bind TRs. Thus our studies have supported the hypothesis that thyroid hormones inhibit α and TSHβ gene expression by the direct binding of a T3-TR complex to specific areas within regulatory regions of the subunit genes and their resultant interference with transcription.

Multiple Thyroid Hormone Receptors

Another interest involves the elucidation of the trans-acting factors involved in thyroid hormone action. Critically important are several forms of the TR. These are encoded by two genes, α and β, each expressing at least two related molecules obtained by alternative splicing. The rat TRα gene encodes a bona fide TR, TRα1, and a related form, c-*erbA*α2, that is identical to TRα1 for the first 370 amino acids, which include the DNA-binding domain and much of the thyroid hormone–binding domain at the carboxyl-terminal end. However, carboxyl terminal to this common region is a divergent sequence. The c-*erbA*α2 form does not bind thyroid hormone or transactivate thyroid hormone–responsive genes, but does bind to putative TREs *in vitro*, as shown by a DNA-binding assay. These and other results suggest that the biologically active c-*erbA*α2 serves as an inhibitor of active TR action. The rat TRβ gene encodes two functional TRs, TRβ1 and TRβ2. TRβ1 and TRβ2 are identical, from several amino acids amino terminal to the DNA-binding region to the remainder of the molecule. However, the amino-terminal regions of the DNA-binding domain are different. We have shown that TRβ2 binds hormone with the same relative affinities as the other receptors, binds DNA, and transactivates thyroid hormone–responsive genes.

Surprisingly, TRβ2 is expressed only in the pituitary gland. This is remarkable 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.

In addition, we wished to determine how these trans-acting factors, of which TRs represent prototypes, are themselves regulated. We have shown that thyroid hormones regulate TR mRNAs in a complex and a tissue-specific manner *in vivo* and *in vitro*. The TRβ2 mRNA is the most regulated form: in a 24-hour period, thyroid hormones decrease TRβ2 mRNA to less than 10 percent of basal levels in the pituitary gland. In contrast, TRα1 and c-*erbA*α2 mRNAs are modestly suppressed, while TRβ1 mRNA is not negatively regulated by thyroid hormones.

In addition, we have observed an interesting feature of the α gene—that the strand opposite the one that encodes TRα1 and c-*erbA*α2 (produced by alternative splicing of a single α transcript) encodes a novel member of the c-*erbA* (thyroid hormone/steroid hormone) supergene family, called *Rev-erbA*α. The predicted protein is more related to thyroid hormone and retinoic acid than the steroid hormone receptors. The mRNA of *Rev-erbA*α shares a common 269–base pair exonic region with that of c-*erbA*α2. This unusual genomic arrangement, which is conserved among several mammalian species, suggests possible interregulation of the TRα and *Rev-erbA*α at the post-transcriptional levels.

Finally, we have recently demonstrated the existence of TRAP (TR auxiliary protein), a nuclear protein that can augment the binding of TRs to various TREs. This ubiquitous 60- to 65-kDa protein dimerizes with TRs to form a heterodimer and also binds specific sequences within the TRE. The full nature of TRAP and its role in thyroid hormone action remains to be determined.

Thus the thyroid hormone receptor family is complex. There are at least three biologically active forms expressed in a tissue-specific fashion and another form that may have an important role in modulating the effects of the others. A knowledge of the functions of these different forms and their interaction with other nuclear factors will be critical for our full understanding of thyroid hormone action.

Summary

Our work has focused on the molecular mechanisms involved in the regulation of TSH gene expression by thyroid hormones, including the cis elements and trans factors that are involved in this process. We hope that our efforts will provide insight into the hormonal regulation of gene expression and cellular function in normal and pathologic 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 Fellow at the University of California, San Francisco.

THE study of the linear sequence of bases in genomic DNA and messenger RNA is steadily gaining momentum 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, yeasts, a plant, and a worm chosen for their well-studied genetics and their small genome sizes. Their genus names are *Mycoplasma*, *Mycobacteria*, *Escherichia*, *Thermococcus*, *Arabidopsis*, and *Caenorhabditis*, and they represent all major branches of the evolutionary tree. The genome closest to completion is that of *Escherichia,* with 30 percent of its 4.7 million base pairs already in the database, through the efforts of around 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 sequencing steps. Because each mixture can be sequenced 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 procedures have been applied to collect over 1 million bases of raw data and are undergoing further development. Chemiluminescent detection of the multiplex sequence images is showing promise as an effective replacement for the radioactivity normally used. To fill in the inevitable last gaps in the sequences, we are exploring several approaches, including direct genomic sequencing (without cloning) from the edges of the gaps and isolation of gap-spanning clones by clone hybridization.

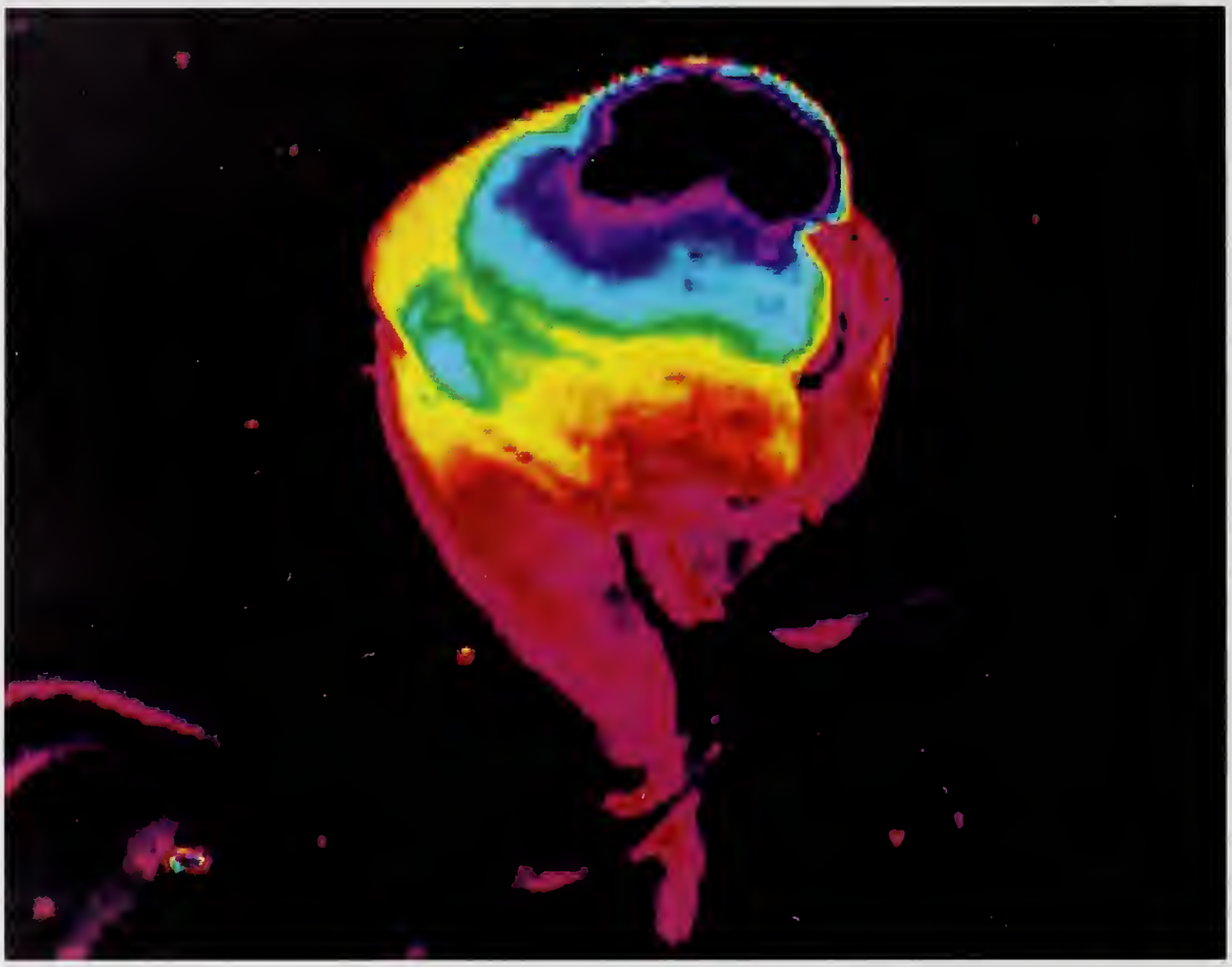
To make the extensive DNA sequences even more useful to biological searches, encoded proteins must be found and collected into families based on their interactions or distant relatives. For example, we have found matches for about half of the genes required by *Escherichia coli* for vitamin B_{12} biosynthesis and have gathered these into families of proteins involved in membrane transport, heme and corrin ring methylation, amine group transfer, and so on.

As another example, we have searched for the class of the most abundant cellular proteins, which have nonetheless eluded the extensive biochemical and genetics studies of *E. 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 130 unique sequences determined so far, over 40 are candidates for such major novel proteins.

We are also extending our methods for detecting *in vivo* molecular interactions by analyzing protection of individual bases from chemical and enzymatic methylation. This *in vivo* footprinting method has been extended to a cloning-based assay for DNA-protein interactions.

In the future, with new sequencing technolo-

gies 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 by investigators worldwide of human sequences and genetics should become more manageable.



Computer-generated false-color image of a Drosophila *leg imaginal disc. The disc was labeled to visualize the mRNA transcript of the* Distal-less *gene. A digitized image of a labeled disc was prepared and processed to generate an intensity "map" so that different colors represent regions of different staining intensity in the original sample. The black region (rimmed in purple) represents the highest RNA concentration. RNA decreases across the disc through several color steps (blue, green, yellow, red, pink), which indicate progressively lower concentrations. The point of this manipulation is to demonstrate that a gradient of the* Distal-less *mRNA exists in the disc. The use of false-color intensity mapping highlights the idea that different spatial domains could in principle be organized or specified by gradients of gene product intensity.*

Research of Stephen M. Cohen.

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-Institut für Entwicklungsbiologie in Tübingen and at the University of Munich in Germany.

ONE of the most striking features of early embryogenesis is the rapidity and reproducibility with which the egg organizes the embryonic body plan. The process by which cells in the embryo acquire distinct identities is known as pattern formation. The concepts are as old as descriptive embryology, and the basic problems are the same in all animal and plant embryos. In order to generate an embryo with an integrated, coherent body pattern, cells need to adopt developmental programs compatible with those of their neighbors. We are interested in understanding how the embryo can orchestrate its genetic information to specify the identities of groups of cells within the context of a larger body pattern. To address these questions, we are analyzing the genetic control of limb development 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. How is the position identified? How do cells that occupy the correct position become committed to develop as limb precursors? Once the limb primordium is established, how do cells get the information that tells them what part of the limb to make?

Specification of Limb Cell Identity

Using classical genetics, we have identified a gene named *Distal-less* that plays an important role in limb development. Limbs do not develop in embryos that lack *Distal-less* gene function. This observation suggests that *Distal-less* activity may play an essential early role in limb development, but it does not tell us what the nature of that role might be. We get an additional hint from examining genetically mosaic animals in which *Distal-less* activity is removed from a group (or clone) of cells that were already committed to develop into limb structures. These cells shift their program and develop into body wall. *Distal-less*, then, seems to function as a developmental switch that is important for telling a cell that it is a limb precursor. *Distal-less* may also define and maintain the identity of the presumptive limb cells.

How might *Distal-less* affect cell identity? To approach an answer in molecular terms, we have cloned and characterized the gene. *Distal-less* is a member of a growing family of important regulatory genes that encode a sequence-specific DNA-binding motif called the homeodomain. A number of *Drosophila* homeodomain-containing genes have been shown to control important developmental decisions by regulating the expression of other genes. We hope to identify the transcriptional targets through which *Distal-less* exerts its regulatory functions.

How Does the Embryo Position Its Limbs?

Using the *Distal-less* gene as a molecular marker, we have been able to visualize populations of limb precursor cells in the embryo well before any overt sign of limb development begins. These small clusters of cells, placed at well-defined sites, are called the limb primordia. Clearly, to define these points, the embryo must use some preexisting source of information, which leads to the spatially localized activation of *Distal-less*. A number of observations implicate the *wingless* gene, without which (in mutant embryos) the limb primordia do not form. The gene is expressed in a reiterated series of well-defined stripes perpendicular to the anterior-posterior axis of the embryo. These and other observations suggest that the *wingless* gene provides a positional cue by which limb formation is oriented.

If *wingless* alone sufficed to define the limb primordia, we would expect the limbs to begin development as stripes of cells expressing *Distal-less*. The limb primordia, however, arise as spots, not as stripes. One way to limit the limb primordia to spots would be to require interaction between the stripes of cells expressing *wingless* and an intersecting longitudinal stripe of cells providing a different type of information. The *wingless* gene encodes a secreted growth-factor-like molecule homologous to the INT-1 oncogene of mammals. A longitudinal stripe of cells

expressing a receptor for the *wingless* signal would allow for specification of clusters of cells at unique points with respect to the body pattern.

Distal-less-Dependent Genes Involved in Limb Development

One approach to identifying genes that act downstream of *Distal-less* in organizing the limb involves looking for other genes expressed in the embryonic limb primordia. We would expect a downstream target gene to depend on *Distal-less* gene function in order to be expressed in the limb primordia. We have begun to screen through the genome, using a method for visualizing the pattern of gene expression rather than for destroying the genes by mutation. Although this work is at an early stage, we have identified a gene, named *disconnected*, that is expressed in the embryonic limb primordia and that depends on *Distal-less*.

Proximal-Distal Pattern Formation

As mentioned above, flies that lack *Distal-less* gene activity do not develop limbs. Flies in which *Distal-less* gene activity is merely impaired develop limbs that are foreshortened along their proximal-distal axis. A range of *Distal-less* mutations of varying severity have been isolated that cause a graded series of defects in limb development. The characteristics of different mutations tell us that the amount of activity of the gene is somehow important in controlling the range of structures that the limb can develop. Distal parts of the limb require more *Distal-less* gene activity than proximal parts in order to develop normally. These observations suggest that the gene may play an important role in organizing the proximal-distal axis of the limbs.

In order to understand how *Distal-less* exerts this function, we have begun to examine the spatial distribution of the gene product in developing limbs of the adult fly. *Distal-less* RNA is expressed in a concentration gradient across the developing limb. The distal-most region of the limb expresses a high level, intermediate regions express lower levels, and proximal regions express little or none. These observations are particularly intriguing in that we have observed a differential requirement for the activity of the gene in different regions of the leg primordium. Distal regions require more activity and express more transcript. We are trying to define better the characteristics of this gradient and hope to determine whether the observed graded distribution of the gene product is contributing meaningful information to the organization of the proximal-distal pattern of the developing limb.

To assess the role of *Distal-less* in organizing pattern in the developing adult limb, we have sought genes that might act downstream of *Distal-less* in this process. The *Abnormal leg pattern* (*Alp*) gene was identified as a prospective partner on the basis of a specific phenotypic interaction with *Distal-less*. The severity of the leg defects caused by dominant mutations of *Alp* is sensitive to changes in the amount of *Distal-less* gene activity. During the past year, we have generated mutants that eliminate *Alp* function and have cloned and characterized the gene. It is expressed in a narrow ring of cells that encircles the developing limb. This ring can be taken as a marker for a defined cell identity along the proximal-distal axis of the leg. Expression of *Alp* in this ring depends on the *Distal-less* gene.

We suggest that understanding pattern formation in the developing limb of the fly will teach us about fundamental mechanisms that may be important during development of vertebrate embryos. The system of pattern formation that we are studying differs from early pattern formation in the precellular fly embryo in that pattern is being organized in a field of cells, rather than in an array of nuclei sharing a common cytoplasm. Cells in the developing limb must define their own identities and communicate this information to their neighbors. In this context we are particularly intrigued by the role that *Distal-less* plays as a pattern organizer at the genetic level and by 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. This year he received the Gairdner Foundation's International Award and shared the National Neurofibromatosis Foundation's Friedrich von Recklinghausen Award.

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.

The most recent edition of *Mendelian Inheritance in Man* lists over 4,000 genetic disorders. In 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 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.

A major development in the past year has been the successful cloning of the gene for von Recklinghausen neurofibromatosis (NF1), a common genetic disorder (sometimes incorrectly referred to as the Elephant Man disease). Using techniques of chromosome jumping and yeast artificial chromosomes, we were able to identify on chromosome 17 a large gene that harbors unmistakable mutations in patients with NF1. Analysis of the gene reveals that it is similar in interesting ways to other genes in humans (and even in yeast). It probably participates in the suppression of unregulated cell growth by interaction with another class of cancer-related genes (the *ras* genes). These developments will make it possible to define in precise terms the basic defect in NF1. In the relatively near future, this gene discovery will lead to improvements in diagnosis; in the longer term it offers the promise of improved therapy for this devastating disease.

In the previous year, 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 an elevation in sweat chloride. Additional investigations of this gene, denoted CFTR (cystic fibrosis transmembrane conductance regulator), has proceeded on many fronts this year.

While one mutation (called ΔF508) is responsible for about 75 percent of CF chromosomes in the United States, we and others have identified numerous mutations in the CFTR gene in other individuals with the disease. We are also studying the regulation of the gene—trying to determine why it is expressed in some tissues (such as lungs and pancreas) but not at all in others (such as brain). It also appears that the gene is capable of producing more than one protein by a phenomenon referred to as alternative splicing.

Importantly, we have succeeded in transferring a normal version of the CFTR gene into CF cells growing in laboratory culture, and have shown that the defect is corrected. These results, achieved in collaboration with James Wilson (HHMI, University of Michigan) and Ray Frizzell (University of Alabama), can be thought of as the first step toward a gene therapy for CF.

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 chromosome jumping

and yeast artificial chromosome cloning from probes known to map to this interval, we have nearly completed an overlapping set of clones of the region and are beginning to search for candidate genes.

A final project is an investigation of the switch in production from fetal to adult hemoglobin in the human. In this precisely controlled phenomenon, fetal hemoglobin genes turn off and adult hemoglobin genes turn on at about the time of birth. This phenomenon is important for two reasons. First, it is a model for developmental control of gene expression. Second, turning the fetal genes back on in an adult with sickle cell anemia or thalassemia would be likely to cure these blood diseases, which afflict hundreds of thousands of persons worldwide.

In attempting to understand this complex process, we have identified a series of mutations in the fetal hemoglobin genes responsible for failure to turn these genes off completely during adult life, which leads to a benign condition called hereditary persistence of fetal hemoglobin (HPFH). These mutations mark DNA sequences important in the developmental switch. We are characterizing proteins from the nucleus that bind to these sequences and control the expression of the hemoglobin genes.

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 explore 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.

The thymus is the source of T cells in all vertebrates, but the central tissue for B cell production may vary. In birds B cells are derived from the hindgut bursa of Fabricius, whereas in mammals they are generated in blood-forming organs, first the fetal liver and then the bone marrow. Multipotent stem cells in these hematopoietic tissues serve as the precursors of T cells, B cells, and other types of blood cells.

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

Genes encoding the T cell antigen receptors (TCR) and antibodies undergo controlled rearrangement and expression in lymphoid cells beginning their development along T or B cell lineages. Other genes encode growth- and differentiation-promoting proteins and their receptors on T and B cells. An elaborate developmental program is thus responsible for the generation of millions of lymphocyte clones, each expressing a TCR or antibody molecule of different antigen specificity. These newly formed T and B cells are seeded via the bloodstream from the thymus or bone marrow to the spleen and other peripheral tissues, where they execute immune surveillance of foreign and self antigens.

T Cell Development

We have embarked on a comparative analysis of T cell development in representative avian, amphibian, and mammalian species, prompted by an interest in the evolutionary strategy for generating T cells that can discriminate between self and nonself and in seeking fresh clues to some of the unresolved mysteries of the human immune system.

Our studies in birds reveal remarkable conservation of the pattern of T cell development found in mammals, including the sequential development of T cells bearing TCR of either $\gamma\delta$ or $\alpha\beta$ isotypes. While the functional role of those expressing $\gamma\delta$ TCR is still enigmatic, they are always generated first during ontogeny, may not undergo clonal selection during their intrathymic development, migrate preferentially to red pulp areas of the spleen and the epithelial lining of intestines, and constitute approximately one-third of the T cell pool in adult birds.

The avian $\gamma\delta$ T cells (also called TCR1 cells) are capable of killing other cells, but unlike the $\alpha\beta$ T cells, are rarely triggered by recognition of conventional class II molecules of the major histocompatibility complex (MHC). Study of their physiological role is facilitated in birds by their relative abundance, and we hope to exploit the experimental ability to inhibit their development in a selective fashion.

A fascinating aspect of the avian $\alpha\beta$ T cells is their development along two discrete sublineages identifiable by the TCR2 and TCR3 monoclonal antibodies. In collaborative studies with Craig Thompson and his colleagues (HHMI, University of Michigan), $TCR2^+$ and $TCR3^+$ cells have been found to utilize different families of TCRβ variable-region genes (Vβ): TCR2 cells use the Vβ1 genes, and the TCR3 cells use Vβ2 genes. The avian Vβ gene locus is thus simpler than the mammalian Vβ locus, which contains many Vβ gene families.

The avian Vβ1 and Vβ2 genes exhibit highly conserved sequences that characterize two major subgroups of the mammalian Vβ genes. Birds may thus provide a useful model system to study the functional significance of the generic Vβ1 and Vβ2 gene families. Avian TCRβ diversity is created largely by nucleotide sequence variations in the joints between the rearranged Vβ, D (diver-

sity), and J (joining) genes. During development, the avian Vβ1 genes are rearranged before the Vβ2 genes, and T cells using the Vβ1 gene family preferentially migrate to the intestines. Notably also, the relative frequencies of alloreactive TCR2 and TCR3 cells vary according to the MHC class II gene combination in a graft-versus-host assay.

Studies of T cell development in the South African frog *Xenopus laevis* are also beginning to reveal interesting information. Monoclonal antibodies have been made that identify $\gamma\delta$ and $\alpha\beta$ TCR candidates expressed as dimeric units by separate lymphocyte subpopulations in the frog thymus, spleen, and intestine. The frog T cells that bear the putative $\gamma\delta$ and $\alpha\beta$ TCR homologues exhibit specialized tissue localization patterns virtually identical to those defined for their avian and mammalian counterparts. Present efforts are directed toward identification of the frog TCR genes and exploitation of the advantages of this amphibian developmental model.

B Cell Development

B cell precursors are identifiable in the bone marrow as lymphocytes that contain μ heavy chains in their cytoplasm. Lacking conventional light chains that are needed for surface expression of immunoglobulins, pre–B cells have been considered "blind" to environmental antigens. This view has been challenged by the recent discovery of genes called λ5 and V pre-B that encode a surrogate light-chain complex apparently expressed together with μ-chains on leukemic pre–B cells. We have identified a subpopulation of normal pre–B cells that express the surrogate light-chain/μ-chain receptor complex on their surface. The data imply that pre–B cells are vulnerable to clonal selection. Current experiments are directed toward identification of the natural ligands for these unusual receptors and characterization of the cellular response.

More than half of the antibodies that we produce daily are of IgA isotype, and many of these are transported across mucosal surfaces of the body for protective immunity. We have identified a receptor for IgA antibodies on phagocytic white blood cells and have characterized its biochemical nature. Monoclonal antibodies made against these IgA receptors are being used to define their precise tissue distribution and for functional studies. These studies focus attention on IgA antibody-mediated phagocytosis as a potentially important defense mechanism.

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 transfer of genetic information from DNA to cell components is the synthesis of an RNA copy of the gene. This process is termed *transcription*, and the RNA synthesized is called a messenger RNA (mRNA). The process can be thought of as making a copy of part of the architectural plan for the construction of a building. The synthesis of each protein component of the cell is directed by a distinct mRNA molecule, and these mRNA molecules are all synthesized by an enzyme called RNA polymerase II. This enzyme recognizes each gene and synthesizes the appropriate amount of mRNA at the required time.

RNA polymerase II contains more than 10 different proteins whose precise functions are largely unknown. Our goal is to understand the structure and function of the subunits of RNA polymerase II. Six years ago my laboratory 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, while absent in bacteria, is found in the 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.

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 phos-

phorylation 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. The identification of a cell cycle control protein as a component of an enzyme that modifies RNA polymerase suggests that cell growth may be controlled, at least in part, through phosphorylation of the CTD.

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.

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. 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 A. 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 the channels that detect sound.

Channels that have been studied in other tissues are of two broad classes: those that are activated by the voltage across the cell membrane, and those that are activated by some chemical either inside or outside the cell. The channels that we study are part of a novel third class. They are directly activated by a mechanical stress on the channel protein, in this case the stress of a sound vibration.

We are especially concerned with those aspects of mechanically sensitive ion channels that give them their unique sensitivity. What are the cellular structures that convey a stimulus to the channel protein, and what stimulus do they convey? What is the biochemical nature of the external link to the channel? What force on the channel protein is required to open it? Are there various conformations of the protein that are closed or open, and what are the energy differences among them?

From our work and that of others, an attractive model has emerged for the way vibration causes a tension on the ion channels in cells of the inner ear. On the top surface of these cells are cilia that rock back and forth with each cycle of a sound wave. Tiny filaments, recently found to run between the tips of the cilia, may connect directly to the channels, such that oscillations of the cilia alternately stretch or relax the filaments, opening the channels or letting them close. High-speed electrical measurements indicate that these channels can open within a few millionths of a second following a mechanical stimulus.

Although the essence of this model remains to be tested, we have worked on some important aspects. We recently confirmed the location of the mechanically sensitive channels at the tips of the cilia, with dyes that change fluorescence when they bind ions coming in through the channels. Measurement of the mechanics of the cilia by high-resolution videomicroscopy indicates that all the channels of a cell receive essentially the same stimulus.

To understand what proteins form the structural links, we must first learn what proteins constitute the cilia. To this end, we developed a method for studying the cilia that utilizes their adhesion to sticky paper and separation of their proteins according to size and electrical charge. We have found about 12–18 different proteins as major constituents and have identified several by size and their reaction with certain antibodies. Much work remains to identify the rest and learn their roles.

In addition to this transduction mechanism, there is an adaptation mechanism in these cells that renders them sensitive to extremely small displacements while they respond over a large range of stimuli. This seems to work by a continuous adjustment system, acting to set the tension on the channels. Within a tenth of a second, the system can restore the resting tension, which is enough to keep some of the channels open at any time.

Our experiments and those of others suggest that the adaptation comes about by a movement of the points where the filaments are attached to the cilia. When the filaments are stretched to open channels, the attachment points slip to allow the fibers to shorten. Conversely, when the filaments are relaxed, the attachment points climb to stretch them, to restore the resting tension on ion channels. The slipping process is accelerated by the movement of calcium into the tips of the cilia, suggesting the mechanism is there. Alteration of the calcium changes the resting tension set by this motor. It also causes a tiny movement of the bundle of cilia—by about 1 millionth of an inch—which we can observe with videomicroscopy. Thus the mechanism

seems more like an active, force-producing motor than a relaxation.

Further evidence from our laboratory suggests that the "motor molecule" that moves the attachment point is like myosin, the protein that causes muscle to contract, and we are now seeking to confirm or refute this analogy.

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. Such a description would 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.

The location of our laboratory in a hospital has also encouraged collaborative projects related to neurological disease. One of the more interesting recent projects involves an inherited muscle disorder, hyperkalemic periodic paralysis. This genetic disease, which causes sporadic weakness or paralysis, is dominantly inherited, so that a child has a 50-50 chance of being affected.

Exercise or certain foods, which raise the level of blood potassium, can bring on a paralytic attack. Earlier experiments had indicated that the increased potassium can change the resting voltage of the muscle and (perhaps indirectly) activate special channels that pass sodium ions. Collaborators at the Massachusetts General Hospital showed that the disease is genetically linked to the muscle sodium channel gene, implicating these channels. We then measured the activity of sodium channels in muscle from an affected patient and found them defective. The defect explains the pathology. Elevated potassium prevents some of the defective sodium channels from closing, and the steady influx of sodium changes the voltage so as to open—and then permanently close—all the other channels. The muscle then cannot contract, causing paralysis. Consistent with dominant inheritance, a small fraction of defective channels indirectly inactivates all the rest, including the normal ones inherited from the other parent. Understanding the pathology at the molecular level gives hints for effective drug treatment.

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 the 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.

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 able to coordinate the actions of other cells involved in the immune response and carry out a variety of immune functions. This pathway of differentiation 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 necessary accessory signals 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 (lasting only about 30 minutes), all of the molecular events required for the decision to proceed down this 10- to 14-day process of cellular differentiation must occur during this short period. Our laboratory is seeking an understanding of the molecular basis of this cellular decision.

By analyzing the regulatory regions of genes that are essential for T cell activation, such as interleukin-2, we have identified several molecules that appear critical to initiating T cell activation. These molecules interact on the promoters of genes necessary for T cell activation in such a way that all must be present for the gene to be activated. Thus the all-or-none decision of a T lymphocyte to differentiate appears related to cooperative interactions between molecules binding to DNA and activating the genes, again in an on-or-off manner, essential for progression toward differentiated function.

Recently, our studies of the immunosuppressive drug cyclosporin A have given us insight into the relative importance of the events involved in the initiation of T cell activation. Although the mechanism of action of this drug is largely unknown, it appears to work early during the commitment period for T cells. By doing this, it blocks the late functions of T cells and also many of the functions of B lymphocytes and other hematopoietic cells that are directed by T cells. One of the proteins we identified in our early studies, nuclear factor of activated T cells (NFAT), is exquisitely sensitive to the effects of cyclosporin A, while nearly all other proteins are unaffected. Other groups had found that cyclosporin A binds and inhibits the function of a cis-trans prolyl isomerase. 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 cyclosporin A and FK-506 (another immunosuppressant) but rather that an inhibitory complex formed between the isomerase and the drug blocks signal transduction. In other studies we found that the specific transcriptional activity of NFAT, but not its DNA-binding activity, is affected by cyclosporin A and FK-506, suggesting

that NFAT might be the direct target of these drugs.

We have found that NFAT is a complex heterodimeric protein: one subunit 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 they do not interfere with the induction of the nuclear component. Thus the prolyl isomerase may function both as an isomerase and as a translocation molecule.

Endodermal and Hepatocyte Differentiation

Several years ago we identified a tissue-specific transcription factor, HNF-1, that interacts with essential regions of the promoters of a large family of genes expressed in endodermally derived tissues. During the past year we were able to purify the protein, obtain the amino acid sequence from protease-derived fragments, and clone its gene. Not surprisingly, the protein contains a homeodomain similar to that found in genes determining body form in insects. Curiously, the protein dimerizes through a region in the amino terminus of the protein, 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 an overlapping group of tissues with HNF-1α, contains a dimerization and homeodomain similar to those of HNF-1α but a different transcriptional activation domain. The HNF-1β pattern of expression indicates that it is a repressor rather than activator of genes selectively expressed in the liver. During the coming year, we will investigate the role of these proteins in mammalian endodermal development.

Regulation of Human Retroviral Gene Expression

Bryan R. Cullen, Ph.D.—Associate Investigator

Dr. Cullen is also Associate Professor in 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 Inc.

RETROVIRUSES derive their name from their ability to reverse the normal flow of genetic information from DNA to RNA. They have the unique 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 transcriptional machinery.

The infection of an animal by retroviruses can result in a number of disease states, of which the most common is leukemia. The avian leukemia virus (ALV), discovered by Ellerman and Bang in 1908, 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.

In the past I worked extensively on the regulation of gene expression by the avian leukemia viruses. This background has greatly facilitated the current research of this laboratory, which focuses on the regulation of gene expression within the human retroviruses and particularly on HIV-1. The genome of HIV-1, like the genomes of the avian and murine leukemia viruses, contains sequences encoding the viral structural genes *env* (envelope glycoprotein), *gag* (capsid protein), and *pol* (RNA-dependent DNA polymerase). However, the complexity of the HIV-1 genome, despite its similar size, is far greater than observed for these animal retroviruses. In particular, HIV-1 is now known to encode six additional gene products that have no equivalents in the avian and murine leukemia viruses. Two of these viral proteins, Tat and Rev, are nuclear regulatory proteins that are essential for HIV-1 replication. Both of these viral trans-activators appear to possess unprecedented mechanisms of action that involve highly specific RNA-protein interactions.

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, particularly, elongation. This mechanism, which appears 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 pro-

teins 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. Recent data suggest 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. Rev function appears to require the sequential binding of multiple Rev protein monomers to the RRE sequence.

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 the interaction with the RRE, while the second appears to interact with a currently unidentified cellular protein that may be 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 transactivator. 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, my laboratory has begun to expand our research to other human retroviruses, including HIV-2 and the apparently nonpathogenic human foamy retrovirus, 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 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 (Mo-MuLV), a member of a group of related leukemogenic retroviruses found in virtually all vertebrates. We have isolated a molecular clone, *Rec-1*, which confers Mo-MuLV infectivity upon introduction into mammalian cells that are not normally susceptible to infection. Subsequent experiments have demonstrated that *Rec-1* encodes for a membrane protein that serves as the Mo-MuLV receptor. Our current research is addressed toward dissecting the molecular details of the virus-receptor interaction that mediates infection.

The Mo-MuLV receptor is not present in mice for the convenience of the virus, but rather must have a function essential to normal cell metabolism. We have identified the receptor protein on the plasma membrane of all mouse cells, suggesting it participates in basic cell metabolism and does not perform a specialized function limited to a subset of tissues. Examination of the amino acid sequence encoded by *Rec-1* reveals that it comprises an abundance of extremely hydrophobic amino acid residues. This implies the receptor protein must exist primarily within the lipid environment of the membrane. A molecular model of the Mo-MuLV receptor predicts that it traverses the plasma membrane as many as 14 times, a structure similar to a protein found in yeast that transports the amino acids arginine and lysine. Arginine and lysine are 2 of the 20 amino acids that are the building blocks of proteins. Frog eggs injected with *Rec-1* RNA demonstrate a large increase in the transport of arginine and lysine across the cell membrane. More detailed experiments have confirmed that the Mo-MuLV receptor is the transporter for all cationic amino acids, i.e., amino acids that carry a net positive charge. Amino acid transporters have been known to exist in mammalian cells, but none have previously been isolated or their structure known. The similarity between the yeast and mouse cationic amino acid transporter is consistent with conservation of a single mechanism for transport of cationic amino acids over evolutionary time and predicts that proteins similar to the Mo-MuLV receptor are used by all animals. Inherited disorders of cationic amino acid transport have been described in patients; mutations in *Rec-1* genes may explain these disorders, a hypothesis we are now examining.

In addition to protein synthesis, cationic amino acids have other roles. For example, arginine and ornithine are important intermediates in the urea cycle, a metabolic pathway found in liver cells that rids animals of nitrogen waste. A protein that is closely related to the Mo-MuLV receptor has been identified in liver tissue, and we are examining its role in transport of ornithine across the mitochondrial membrane of hepatocytes, an important step in this pathway. Also within the past few years, arginine has been identified as the substrate for nitric oxide, the proximal molecule in a cell-signaling pathway that is important in the control of blood pressure, nerve transmission, and host immune defense. Currently we are investigating how the Mo-MuLV receptor/transporter can influence nitric oxide production by regulating arginine availability.

The laboratory remains interested in how retroviruses interact with the Mo-MuLV receptor to permit fusion with the target cell membrane and

gain entry into the cell. Recently we completed experiments that identify the portion of the Mo-MuLV receptor that binds to the virus envelope. Now we are examining other mutant receptor proteins that have normal retrovirus binding but do not permit infection. These studies may identify portions of the receptor protein that are important for the fusion of the virus to the target cell, the step in infection that follows binding. We are also investigating the consequences of Mo-MuLV binding and infection on cationic amino acid transport. Our long-term goal is to understand the chemical basis of the Mo-MuLV–*Rec-1* interaction in sufficient molecular detail to design small molecules that can block virus binding and prevent infection.

Mechanisms of Antigen Presentation

C. Geoffrey Davis, Ph.D.—Assistant Investigator

Dr. Davis is also Assistant Professor of Medicine and of Microbiology and Immunology at the University of California, San Francisco. He received his B.A. degree in biology from Swarthmore College and his Ph.D. degree in immunology from UCSF. He did his postdoctoral research in the laboratory of Joseph Goldstein and Michael Brown at the University of Texas Health Science Center at Dallas.

THE two key features of the immune system are diversity and specificity. While it must be capable of recognizing a large array of pathogens, it must subsequently mount a specific and appropriate response to each. Diversity is accomplished at the level of the cell surface receptors on T and B lymphocytes. Specificity is accomplished at several levels, the first of which is in the intracellular proteolytic processing of the pathogen. This processing results in the generation of peptide fragments, which then bind to either class I or class II products of the major histocompatibility complex (MHC).

The choice of class I or class II MHC molecule is determined by the nature of the pathogen. In general, peptide fragments derived from pathogens that have entered the cell from outside, like most bacteria, bind to class II molecules, while fragments derived from pathogens that use the cell's replicative machinery to proliferate, like viruses, bind to class I molecules. Once bound, the fragment is transported to the cell surface, where it can be recognized by an effector T lymphocyte.

The type of T lymphocyte responding is closely linked to the type of MHC molecule presenting the peptide. Thus a peptide bound to a class II molecule will stimulate a helper T cell, which in turn will trigger the production of antibodies. In contrast, a peptide bound to a class I molecule targets the presenting cell for destruction by stimulating a cytotoxic T lymphocyte. Our research focuses on various aspects of the intracellular processes that allow class I MHC molecules to distinguish endogenously synthesized pathogens from those that have been taken up from outside.

The mature class I molecule is a heterodimer consisting of a transmembrane heavy chain noncovalently associated with a soluble light chain. Studies with mutant cell lines have revealed that the assembly of class I heterodimers takes place in the endoplasmic reticulum and is tightly regulated. These studies have shown that the peptide itself plays a role in determining the fate of the complex. Thus class I molecules that have successfully bound a peptide are transported to the cell surface, while those that have failed to bind a peptide dissociate and are degraded within the cell. This implies that there is an intracellular "gate" that distinguishes heterodimers that have bound peptide—and are therefore functional—from those that have not. To gain insight into the mechanism behind this gate, Rebecca Elstrom has studied the assembly process in an *in vitro* system. She has succeeded in inducing class I heavy and light chains, which have been synthesized *in vitro*, to assemble within membrane vesicles derived from the endoplasmic reticulum. The results of her studies indicate that in an oxidizing environment, which is the physiological state of the endoplasmic reticulum, stable heterodimers are formed even in the absence of peptide. In contrast, in a reducing environment, stable heterodimers are formed only when peptide is added. These findings suggest that class I heterodimers form in the endoplasmic reticulum independent of peptide. Subsequently, while en route to the surface, they may encounter an intracellular compartment with a reducing environment. In this compartment, only heterodimers that have bound peptide remain intact and are allowed to proceed to the cell surface, while those that have not bound peptide dissociate. This *in vitro* system holds great promise for producing further insights into assembly of class I complexes.

Studies initiated by Michael Zegans, a Howard Hughes Medical Scholar, and being pursued by Rabin Chakrabarti, a Howard Hughes Associate, focus on a special problem in antigen presentation, that of intracellular parasites. Such parasites replicate within the host cell but use their own machinery to do so. Although they might be expected to shed proteins into the cytoplasm where they would be exposed to proteolysis and thus to presentation by class I molecules, the occurrence of class I–restricted immune responses has, for many parasites, been difficult to demonstrate. *Chlamydia trachomatis*, a leading cause of blindness in the world, is an organism particularly suitable for such studies, since more than half of its entire mass is contributed by a single

protein—the major outer membrane protein (MOMP). Dr. Chakrabarti has succeeded in establishing stable cell lines expressing the MOMP gene. These cell lines are being used for the detection and characterization of cytotoxic responses in infected mice.

Juerg Baenziger, a Howard Hughes Associate, has been studying another aspect of class I–mediated processes. This study is related to the behavior of the peptide–class I complexes in T lymphocytes after they reach the surface. Several years ago it was observed that class I complexes are endocytosed, or internalized, by T lymphocytes, but only when the lymphocytes are activated. To explore further the molecular basis as well as the function of T cell–specific endocytosis of class I complexes, he has constructed a series of site-directed mutants of the human class I MHC molecule A2.1. These constructs include various truncations, deletions, and substitutions within the cytoplasmic domain of this membrane-spanning protein. They also include chimeras in which the extracellular domain is derived from A2.1 and the transmembrane and cytoplasmic domains are derived from either the T cell surface marker, CD4, or the receptor for low-density lipoprotein (LDL).

Several of the constructs have been introduced into Jurkat cells, a human T cell lymphoma line. The properties of the chimeras in these cells with regard to endocytosis exactly parallel those of the donors of the cytoplasmic tails. Specifically, CD4 chimeras endocytose only when phosphorylated, whereas LDL receptor chimeras are endocytosed and recycled constitutively.

In addition to providing insights into structure-function relationships in the trafficking of the native molecules, studies in these cell lines have led Dr. Baenziger to conclude that the class I molecules that are endocytosed in activated T lymphocytes are structurally distinct from normal class I molecules. By comparing messenger RNA from resting and activated human T cells, he found that activated cells produce a single new species of class I heavy chains. These heavy chains, which lack the domain that normally anchors them in the membrane, are secreted from the cell. Current studies are directed toward determining whether these soluble class I molecules account in whole or in part for the endocytosis observed in activated T lymphocytes. Such a mechanism would have novel implications for regulation of T cell–mediated immune responses.

Molecular Approaches to Lymphocyte Recognition and Differentiation

Mark M. Davis, Ph.D.—Associate Investigator

Dr. Davis is also Associate 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 the NIH as a postdoctoral fellow and Staff Fellow before joining the faculty at Stanford.

WE have focused on several major areas in immunology that generally resolve into three problems: How do T cells recognize foreign entities? How are they selected in the thymus? How is lymphocyte differentiation controlled genetically? An additional goal is to refine and better integrate recombinant DNA technology with some of the other powerful techniques in immunology, as an approach to defining the function of unknown genes or poorly understood genes and their products.

T Cell Recognition of Foreign Antigens

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. This is in contrast to antibodies, which, while closely related to T cell antigen receptors, bind intact foreign proteins directly. Because of the consistently high concentration of sequence diversity in the V-J junctional region of T cell receptors (and other 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 possibility and learn about the molecular dynamics of T cell receptor–peptide–MHC interactions in general, we have developed expression systems that allow us to produce either T cell receptor or MHC class II heterodimers in a solubilizable 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 T cell receptor and its cognate MHC molecule. We hope to demonstrate a direct interaction between these macromolecules (with the appropriate peptide in the MHC component) and to measure the affinity with which this happens. This work is also intended to provide the raw material needed for structural studies, such as x-ray crystallography.

We have already 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), which recycles through low pH endosomal compartments. Our current data suggest that the low pH triggers a specific comformational change in the MHC molecule, which allows it to bind new peptides more easily.

Recent experiments relevant to the issue of whether the CDR3-equivalent regions of T cell receptors are responsible for peptide recognition take advantage of an immunological version of classical genetic second-site suppression analysis. In our 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 T cell receptor sequences. To hold part of the original receptor constant, we immunize mice that are transgenic for either chain of the original T cell receptor. With this new approach, we have shown that of the two residues on the peptide that are most important in the T cell response, one is governed by the CDR3 of the V_α polypeptide and the other, three amino acids downstream on the peptide sequence, is specified by the CDR3 of the V_β polypeptide. Thus we have generated significant support for the original hypothesis.

T Cell Receptor Selection

Another area of interest in T cell recognition involves the ways in which an organism might select for particular T cell receptor–bearing cells. The specific mechanisms and extent of such selections are just now becoming clear. To address these questions, we have made mice

transgenic for T cell receptors of defined specificity and have followed the expression of these molecules during T cell differentiation. We have obtained evidence of selection both for and against particular T cell receptors, depending on the MHC genotype of the host.

Our evidence concerning positive selection indicates that this is an inefficient process that can be significantly influenced by the density or location of the MHC molecules required. Experiments also indicate that MHC alleles that differ in the binding site of MHC profoundly alter or abolish this type of selection, suggesting that specific peptides (or other molecules occupying the site) may play a role. Thus it seems probable that in addition to the complex of foreign antigen plus MHC that a given T cell receptor may encounter in the periphery, there is first another MHC + "X" complex that must be successfully encountered in the thymus.

We have also developed a transgenic model for self-tolerance of T cells by introducing a fusion protein that contains the target peptide seen by our T cell receptor transgenic mice. This, together with direct peptide injection/slow delivery systems, has enabled us to control the timing and dose of an antigen seen as "self." We find evidence for both paralysis and deletion of T cells, with deletion, which occurs in both the thymus and the periphery, the favored outcome as the dose of self antigen is increased.

Genes Controlling Lymphocyte Differentiation

The last area of interest involves the isolation and characterization of genes that might control differentiation in lymphocytes. We have worked extensively on the XLR gene, which encodes a small (25-kDa) previously unidentified nuclear protein that is specifically turned on in late-stage B cells and medium- to later-stage T cells. One novel feature of the protein is that it is stabilized in its nuclear location by a relatively low concentration of zinc ions. A number of other proteins have the same characteristics, and there may be some novel zinc-dependent structures in the nucleus. We are currently trying to deduce what the product of this gene does.

A major effort is also under way to isolate other potential regulatory genes, using new subtractive hybridization and cloning schemes and nuclear localization as a screen. We are particularly interested in genes that are turned on late in B cell differentiation and in those that are specific for early thymocytes. Nuclear localizing proteins in either category may provide important clues about the regulation of differentiation in these cells, clues that may be applicable to cellular differentiation in general.

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 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 factors, 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 the growth factor 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.

Signaling by the EGF Receptor

A principal question that we must answer in order 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 have recently identified a novel protein kinase activity that is acutely regulated by growth factors. This protein kinase is located in both the nuclear and cytosolic compartments of cells. Substrates 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 structure of this protein kinase by molecular cloning, and examining the role of this enzyme during signal transduction.

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.

Three-Dimensional Structures of Biological Macromolecules

Johann Deisenhofer, Ph.D.—Investigator

Dr. Deisenhofer is also Regental Professor of Biochemistry and 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 many honors for his structure analysis of a photosynthetic reaction center, including the 1988 Nobel Prize in chemistry (with Hartmut Michel and Robert Huber).

WE determine and study the three-dimensional structures of proteins in order 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. Some of the most fascinating members of this group are membrane-spanning and membrane-associated proteins.

One of the major obstacles on the way to a structure analysis by x-ray diffraction is the necessity to grow big, well-ordered single crystals from pure samples of the macromolecule under investigation. Although numerous proteins have been crystallized, the crystallization process is still not well understood. Successful crystallizations can provide guidelines, but every different protein presents a completely new problem, and there is no guarantee that good crystals will grow.

This is especially true for proteins that are embedded in biological membranes. For crystallization they have to be removed from their natural environment with detergents. The choice of detergent is critical; moreover, the most successful crystallizations of membrane proteins were performed in the presence of a detergent mixed with a so-called small amphiphile. Thus, when trying to crystallize membrane proteins, we have to manipulate a complex mixture of aqueous buffer, detergents, small amphiphiles, precipitant, and our protein. This complexity is one of the reasons the three-dimensional analysis of membrane proteins is progressing so slowly.

Cytochrome bc_1 Complexes

The chemical reactions during photosynthesis can be divided into "light" and "dark," with the former reactions capturing light energy and storing it, mainly in the form of a proton gradient across a photosynthetic membrane. This gradient is then used for making, e.g., ATP, the universal "energy currency" of living cells.

The photosynthetic light reactions in purple bacteria require three types of molecules: photosynthetic reaction centers (RCs), cytochrome bc_1 complexes, and cytochrome c_2. The first two are integral membrane proteins, and the third is a small soluble electron-transport protein.

The RCs perform the first electron transfer step of the light reactions. A pair of chlorophyll molecules absorb a photon and subsequently release one of its electrons, which moves within the RC through the membrane to a quinone molecule. This electron transfer process must happen twice in succession so that the quinone receives two electrons. The charged quinone picks up two protons and dissociates from the RC as a quinol, which migrates along the membrane to the cytochrome bc_1 complex.

This complex removes the electrons and protons from the quinol and transfers them, with two additional protons from the cytoplasm, back through the membrane. The electrons are shuttled back to the RC, where they can again enter the cyclic light-driven transfer process; the protons accumulate on the outside of the membrane, building up the desired proton gradient. The structure analysis at 2.3 Å resolution of the RC from the purple bacterium *Rhodopseudomonas viridis* has contributed significantly to our understanding of the first half of this cyclic process. Determination of the three-dimensional structure of a cytochrome bc_1 complex would similarly elucidate the second half.

Complexes of the bc_1 type play a crucial role not only in photosynthesis but also in cell respiration. For example, the mitochondria of all higher organisms have such complexes in their inner membrane. The photosynthetic purple bacteria have the simplest bc_1 complexes, consisting of only three different protein subunits, with three heme groups and an iron-sulfur cluster as cofactors. They occur in the bacterial inner membrane at a concentration significantly lower than that of RCs, and are therefore more difficult to isolate. The research group of David Knaff at Texas Technical University, Lubbock, recently succeeded in isolating the bc_1 complex from *Rhodospirillum*

rubrum. In collaboration with Knaff's group, we prepared this complex to high purity and in sufficient quantity to start crystallization experiments, which are now in progress.

DNA Photolyase

Light energy plays an important role in reactions other than photosynthesis. An example of the use of light energy to drive a chemical reaction is found in the molecular machinery that enables cells to repair DNA damaged by ultraviolet light. One type of damage frequently caused by ultraviolet irradiation is the crosslinking of two neighboring thymine bases in a DNA strand. Most of these crosslinks are in the form of a cyclobutane ring connecting four carbon atoms, two from each thymine base. An enzyme, DNA photolyase, can locate and bind to such lesions and, upon input of light of suitable wavelength, cleave the carbon-carbon bonds between the bases, repairing the damage.

DNA photolyase has been found in prokaryotes, eukaryotes, and archaebacteria. The enzyme from *Escherichia coli* has been sequenced, overexpressed, and purified in the laboratory of Aziz Sancar at the University of North Carolina, Chapel Hill. 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's role appears to be that of a light-harvesting antenna.

Despite significant problems with the enzyme's tendency to denature, we were able to crystallize DNA photolyase from *E. coli*. The crystals, obtained after repeated seeding of crystallization solutions, are plates less than 0.05 mm thick, but they diffract x-rays to at least 2.8 Å resolution. That should suffice to construct an atomic model of this interesting enzyme after the phase problem has been solved. Collection of x-ray diffraction intensity data and experiments to bind heavy atom compounds to the protein in the crystal are under way. We also are trying to crystallize the enzyme in complex with a substrate, a five-nucleotide piece of single-stranded DNA containing 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.

Since light is an essential ingredient of the enzymatic reaction of DNA photolyase, the enzyme-substrate complex will be a suitable system in which to study the time course of the reaction, using the Laue technique. This utilizes the broad spectrum of x-ray wavelengths in a 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. Data from such an experiment can provide snapshots of the structural rearrangements during the reaction, thus contributing to the understanding of the enzyme's mechanism.

Other Projects

In addition to the two projects described above, we are doing purification and crystallization experiments on several other proteins. The subjects of these studies include 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; and mammalian phosphofructokinase.

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 B.A. 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. His postdoctoral fellowship was done with David Baltimore at the Massachusetts Institute of Technology.

ONE remarkable feature of the immune system is its ability to recognize and respond to an extraordinarily large variety of foreign molecules, or antigens. This exquisite specificity is mediated through protein receptors that bind tightly to specific antigens. Such receptors are found on the surfaces of two types of immune cells: B cells and T cells.

Antibody molecules, or immunoglobulins, represent one type of antigen receptor. 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; during the maturation of antibody-producing cells (B cells), segments are joined to form intact immunoglobulin genes. In addition to antibody, there is another class of antigen receptor that 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, these receptor molecules are expressed at the surfaces of B and T cells, where specific interactions between receptor and antigen trigger division and further maturation of B and T cells.

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.

Antibody Gene Rearrangement

The rearrangement of antibody genes does not occur haphazardly during development, but is a well-orchestrated process in which some segments are joined first and others later. The available evidence suggests that this orchestration is accomplished by mechanisms related to those that regulate gene expression. Antibodies are made of two kinds of protein chains—heavy and light. The genes that encode the heavy chains are split into three segments—V, D, and J—that are brought together by two rearrangements. First a D segment joins to a J segment; then a V segment is fused to the DJ element.

It has been known for several years that incompletely assembled antibody genes, the DJ elements, are transcribed into RNA. We have recently worked out some of the details of this process. Each D segment carries a DNA sequence that supports initiation of transcription, a promoter sequence. The promoter is inactive, however, unless it is near a second regulatory element, an enhancer. When a D segment joins to a J segment, the D promoter is brought near the enhancer region, and transcription is initiated. The promoters that drive expression of completely assembled antibody genes—the promoters that lie upstream of V segments—are activated in much the same way, but the V and D promoters are structurally quite dissimilar. Thus, at the level of DNA-protein interactions, the D promoters are likely to be regulated differently from the V promoters. An understanding of these differences will shed light on regulatory mechanisms that are in force during the earliest stages of B cell development.

Although specific DNA rearrangements play a central role in the development of the immune system, we know little about the mechanics of this process. We have taken two approaches to the problem. One has been to analyze the products of rearrangement and the DNA sequences that mediate the reaction. The second has been to determine how these DNA sequences are recognized by the machinery that carries out recombination.

To facilitate our study of antibody gene rearrangement, we have made artificial DNA molecules that contain unrecombined antibody gene segments. By simple tests, we can detect rearrangements of these artificial substrates after they are introduced into immature B cells. By following the fates of these engineered molecules we have been able to infer some general features of

antibody gene assembly. Rearrangement of antibody and T cell receptor genes is mediated by conserved DNA sequences (signal sequences) that lie next to the gene segments, near the sites of recombination. It appears that during the initial stages of rearrangement, two antibody segments are paired and then cleaved to form an intermediate in which four DNA ends (the ends of the gene segments and the ends of their flanking signal sequences) are held near each other by the recombinational machinery. One of our aims is to test this idea by recovering such intermediates from B cells that carry model recombinational substrates.

Essential to an understanding of B and T cell development is a molecular description of antibody and T cell receptor gene rearrangement. Our approach to this problem began with the guess that the recombinational signal sequences represent parts of a scaffold on which the recombinational machinery is assembled. We examined extracts of immature B and T cells for proteins that could bind specifically to these signals. Our search uncovered a specific DNA-binding protein whose target coincides precisely with one of the recombinational signals: a conserved 9–base pair (bp) DNA segment that is needed for efficient rearrangement of antibody genes. At the same time, other laboratories identified at least three genes whose products likely play a role in rearrangement. We are now examining the relationship between these gene products and the DNA-binding protein that we have purified, with the goal of understanding what the recombinational machinery is and how it works.

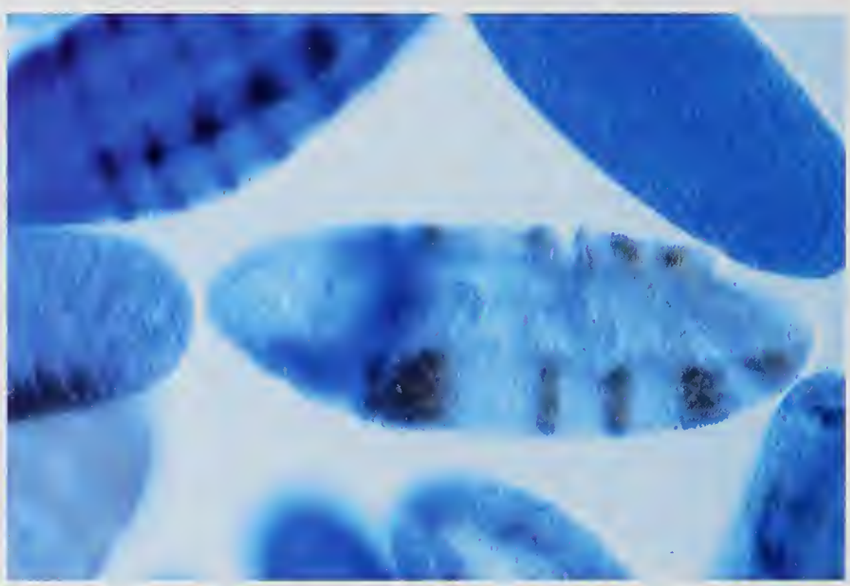
Lymphocyte Activation

A second area of work in the laboratory addresses a different problem: the processes by which specific antigen and growth factors trigger the activation of B and T cells. We have set out to identify molecules that participate in these processes, with the goal of understanding how these molecules function in the generation of an immune response. We have begun by focusing on a group of enzymes called tyrosine kinases. Biochemical evidence has long suggested that these enzymes might be intimately involved in B and T cell activation, but the molecules themselves have proven elusive. Among genes expressed in interacting B and T cells we have found one that encodes a novel tyrosine kinase that is homologous to the cellular proto-oncogene c-*src*. Unlike c-*src*, however, this gene shows a strikingly restricted pattern of expression, being transcribed preferentially in B cells and in developmental precursors of B cells. We have accordingly called this gene *blk* (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 cells or plasma cells, they no longer express these signal transmission proteins and also stop expressing *blk*. This 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.

How might extracellular signals activate the *blk* kinase? Our recent results suggest that the *blk* kinase is inactivated by addition of a chemical group to a specific site on the kinase molecule; the enzyme is activated when this chemical group 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. Expression of this mutant kinase in certain types of cells leads to alterations in growth. By concentrating on the specific extracellular signals that turn the *blk* kinase on, we hope to identify the pathways in which *blk* functions.

Aside from its likely role in B cell signaling, *blk* provides a window into how gene expression is restricted to specific cell types during development. To date, most work on regulated expression in B cells has focused on the immunoglobulin genes. The pattern of expression of *blk* among B cell lines differs from the pattern of expression of immunoglobulin genes. The *blk* gene occupies approximately 30 kbp of DNA on mouse chromosome 14. Its unusual structure carries two transcriptional promoters. One of these is preferentially used early in B cell development, while both promoters are used at later stages. We expect that this temporal regulation reflects differences in the arrays of DNA-protein complexes that drive transcription of *blk* in immature and mature B cells, and we now wish to explore these differences in detail.

1



2

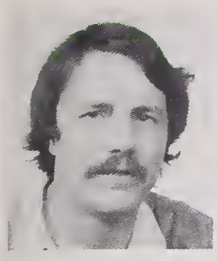
Double staining of wild-type and hunchback-*mutant embryos, using an anti–Fushi tarazu (Ftz) antibody and a DGG-labeled probe for* lacZ *mRNA.*

1. The wild-type embryo exhibits a normal seven-stripe pattern of Ftz expression and shows the expression of a reporter lacZ *gene in the anterior part of the embryo. This reporter gene is controlled by the hsp70 TATA box promoter containing artificial binding sites for both the* bicoid *and* hunchback *gene products.*

2. In the hunchback-*mutant embryo, the pattern of* fushi tarazu *is disrupted in the anterior part, while the reporter is not affected. As discussed in the text, this points to a synergistic activation of the reporter gene by the product of the maternal gene* bicoid *and the maternal component of* hunchback.

Research of Claude Desplan.

Transcription Control During Early Drosophila *Development*



Claude Desplan, Ph.D.—Assistant Investigator

Dr. Desplan is also Assistant Professor and University Fellow 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 that contain conserved domains such as homeodomains or zinc finger domains. We believe the zygotic genome responds to maternal organizing factors through a network of transcriptional regulators to set up the body pattern of the embryo. Since several genes expressed during mammalian embryogenesis share homology with developmental genes of the fly, it is likely that the mechanisms uncovered are of general significance for the development of multicellular organisms.

The goal of our laboratory is to understand the mechanisms involved in these regulatory interactions. 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 of regulatory interactions consistent with properties uncovered through the first approach. We are investigating the mechanism of action of two classes of developmental gene products containing either a homeodomain (HD) or zinc finger DNA-binding motifs. These motifs are known to bind to DNA and to target regulatory proteins to their downstream genes.

Interactions Among Gap Genes

The first zygotic genes to be expressed in the *Drosophila* embryo are the gap genes. Their role is to read and interpret coarse positional information deposited in the egg by the mother and to refine this information by cross-regulatory interactions. Three gap genes, *Krüppel* (*Kr*), *hunchback* (*hb*), and *knirps* (*kni*), have been analyzed molecularly. All contain zinc finger motifs characteristic of many eukaryotic DNA-binding proteins. This homology suggests that they act as regulators of transcription.

A relatively simple circuit appears to control expression of *hb*. It is activated in response to the homeodomain-containing product of the maternal gene *bicoid* (*bcd*) and is repressed by the neighboring gap gene *Kr*. The *hb* promoter contains DNA-binding sites for the products of the *bcd*, *Kr*, and *hb* genes (Bcd, Kr, and Hb), which have been defined *in vitro*. Another zygotic gene involved early in the development of the head is *orthodenticle* (*otd*). Its product (Otd) contains an HD that recognizes the same sequence as the Bcd protein and may be part of the same regulatory network in the anterior part of the embryo.

Hb, Kr, Otd, and Bcd act very early, when the fly embryo is still a syncytium and when few other proteins are expressed. To study the molecular interactions among the genes of origin, we have introduced into transgenic flies artificial genes containing combinations of the binding sites for Kr, Hb, or Bcd/Otd in promoters driving the reporter gene β-galactosidase. The patterned expression of the reporter gene leads to the conclusion that Bcd is not sufficient for expression in the anterior part of the embryo, despite strong Bcd-binding sites. Addition of Hb-binding sites is required to obtain a promoter whose expression is maintained in the anterior part, mimicking that of *hb*. Our work points to a synergistic activation of the promoter by Bcd and the maternal component of *hb*.

The DNA-binding Specificity of the Homeodomain

Over 20 genes involved in development encode HD-containing proteins. The HD includes a domain similar to the helix-turn-helix motif of many prokaryotic DNA-binding proteins. Our analysis of the function of the HD has led us to describe how we can change the DNA-binding specificity of one HD to another, simply by replacing a single amino acid at the critical position 9 of the recognition helix. As we proposed, structural analysis of HD-DNA complexes later showed that the interaction between DNA and the HD is not analogous to that of the prokaryotic helix-turn-helix proteins, but instead 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 recognized.

The Paired Gene Encodes a Multifunctional Transcription Factor

The product of the pair-rule gene *paired* (Prd) contains not only an HD but also other domains conserved in evolution that are found in developmental genes. We have shown that the 128–amino acid Paired domain is also DNA binding, making the Prd protein a bifunctional DNA-binding transcription factor. Although both the Paired domain and the HD can bind to DNA independently, the promoter of another pair-rule gene has revealed sites that are bound cooperatively by Prd when both domains are present in the same protein. Extensive mutagenesis analysis of the protein has allowed the definition of at least four subfunctions in Prd.

Transcriptional Functions of the Homeodomain Proteins Fushi tarazu and Engrailed

In order to understand the mechanism by which proteins with the same DNA-binding specificity can have different developmental functions, we have analyzed the roles of two proteins binding to the same target site in an *in vitro* transcription system. We have shown that the Fushi tarazu (Ftz) protein activates *in vitro* transcription in a binding site–dependent manner. This activation is prevented by the addition of Engrailed (En), which competes with Ftz for binding to the same sites. When more En protein is added, a further repression is mediated by En binding to the basal promoter. We have shown that, in this case, En acts by binding to the TATA box and competing with the TATA box–binding factor TFIID. Formation of a committed complex by preincubation of the promoter with TFIID prevents both the repression by En and its binding to the TATA box.

Targets of the Homeodomain Developmental Gene *engrailed*

The HD transcription factor Engrailed belongs to a small set of genes required for pattern formation within each embryonic segment and within imaginal discs. We are using the "enhancer trap" technique and a +/− screen for cDNA from undifferentiated imaginal disc cells in culture to identify target genes affected by *engrailed*. In this approach, we want to obtain the effector genes essential to the definition of cellular and compartmental identity in discs. These genes may prove to be components of signaling pathways relevant to both early and late development.

We have investigated both the general transcription functions carried out by the HD and zinc finger proteins and the particular properties of each gene product. Our attention is now focused on the combinatorial effects of such factors *in vitro* and *in vivo*, and we expect to acquire insight into the mechanisms controlling the coordinate expression of developmental genes.

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. 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, California. He has received the Iowa Governor's Science Medal and the Burroughs-Wellcome Award in Molecular Parasitology.

MORE than a billion people in tropical areas of the world endure parasitic infections during most of their life. The various protozoan and helminthic parasites that cause these infectious diseases have evolved 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 ways in which to combat or prevent the infections. The parasites that cause three of these tropical diseases are described here.

Leishmaniasis

Leishmania parasites are protozoan pathogens found in many tropical countries, where they cause a spectrum of diseases that include cutaneous, mucocutaneous, and visceral leishmaniasis. The parasites have a two-stage life cycle that occurs in the sandfly vector and a mammalian host. Within sandflies they exist as uniflagellar promastigotes and 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, promastigotes are phagocytosed by host macrophages, where they reside as spherical amastigotes in 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. We are studying how these organisms are able to 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 probably participates in the uptake of the promastigotes by macrophages and contributes to the survival of the amastigotes 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 recently discovered that three different RNA species encoding gp63 occur during cultured promastigote growth. One RNA species is present in the organisms only during the early, less infectious stages of cultured growth, when little gp63 is on the promastigote surface. This gp63 RNA disappears and another gp63 RNA species appears as the promastigotes become more infectious. The amount 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. The three RNAs encode similar gp63 proteins and differ primarily in the nucleotide sequences of their 3′-untranslated regions. They are derived from different genes, and we are determining the molecular mechanisms that regulate their differential expression. 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 gp63 proteases during the promastigote and amastigote stages.

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

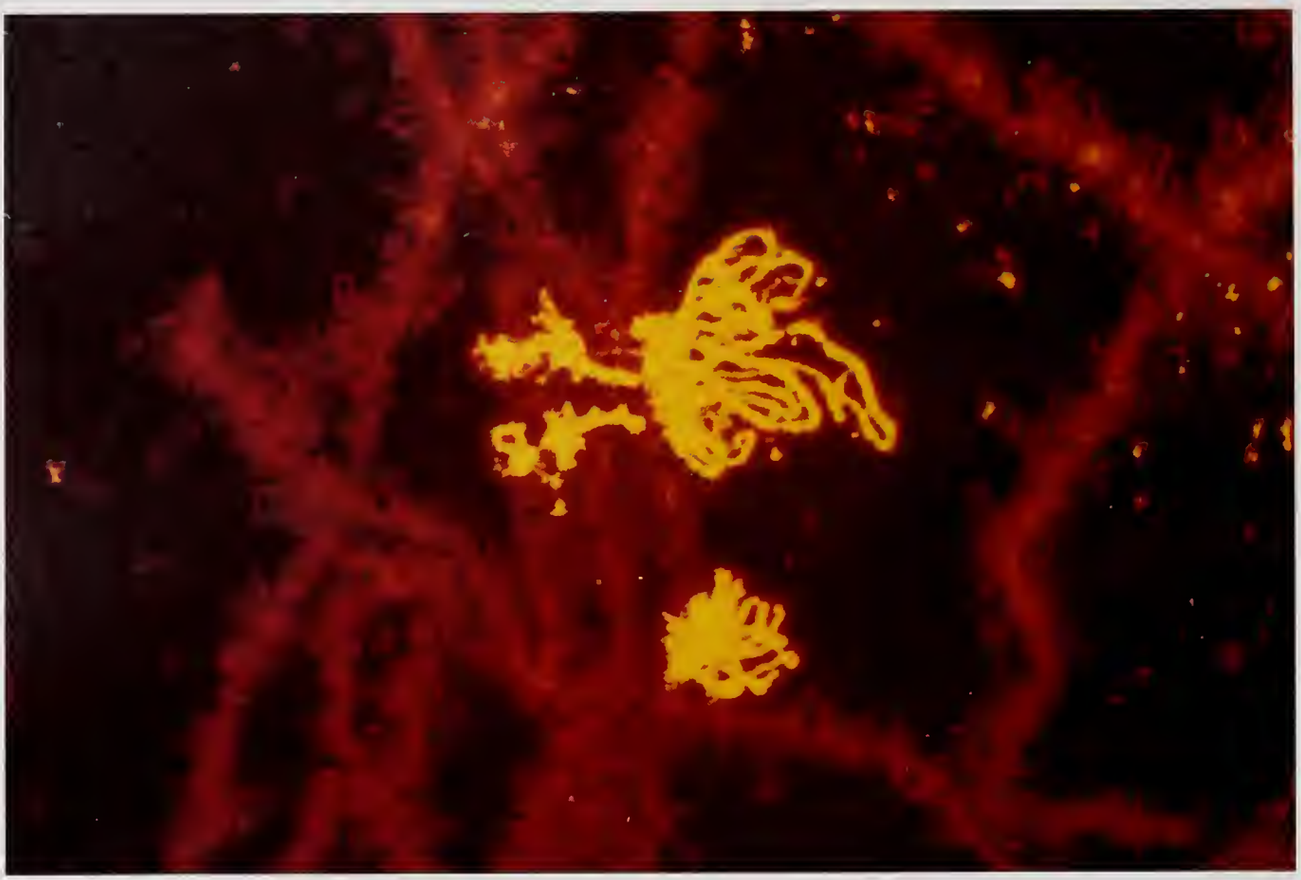
ahead'' of the host immune defenses. We are examining the events at the DNA and RNA level that are responsible for this antigenic variation.

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 the VSG gene 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 adjacent to the ends of the chromosomes—the chromosomal telomeres. The process is complicated by the fact that several, and perhaps many, potential 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. 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 characterization of a group of small chromosomes that are unique to African trypanosomes and contain many of the VSG genes destined for sequential expression.

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. Here 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 in the host evade the immune response. 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 from the L4 stage larvae that quickly develop after infection, and are using specific cDNA clones to overproduce parasite antigens from both stages 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.



Localization by immunofluorescence microscopy of the hnRNP L protein on the lampbrush chromosomes of the North American newt Notophthalmus viridescens. *The intense yellow staining represents a high concentration of the L protein on the landmark giant loops. This is the first protein localized to these mysterious segments of the chromosome. The immunofluorescence was performed with the monoclonal antibody 4D11, in collaboration with Joe Gall of the Carnegie Institution.*

From Piñol-Roma, S., Swanson, M.S., Gall, J.G., and Dreyfuss, G. 1989. J Cell Biol *109: 2575–2587.*

Post-transcriptional Regulation of Gene Expression, Ribonucleoprotein 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 was a Helen Hay Whitney postdoctoral 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), the functional translatable intermediates of gene expression, 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), which describes their variable size and cellular localization. It is possible that only a subset of hnRNAs are actually precursors of mRNA and that the rest turn over in the nucleus.

The collective term for the proteins that bind hnRNAs, and that are not stable components of other classes of ribonucleoprotein (RNP) complexes, is hnRNP proteins. They are significant in that they are bound to hnRNAs and thus influence their structure, fate, and processing into mRNAs. The hnRNP proteins are also as abundant in growing vertebrate cells as histones, and their complexes with hnRNA are also of interest as major nuclear structures.

Once formed, the mRNAs are transported to the cytoplasm via nuclear pores—a little-known process that appears to be one of the most important regulatory steps in the post-transcriptional pathway of gene expression. In the cytoplasm, mRNP proteins are likely to be involved in regulating the translation, stability, and localization of mRNAs. 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 do so we investigate the structure, function, and localization of the hnRNP and mRNP proteins and the RNP complexes.

We have identified the hnRNP and mRNP proteins by photochemical RNA-protein crosslinking in intact cells and have produced monoclonal antibodies to many of them. The antibodies were used to develop an immunopurification procedure for hnRNA-hnRNP complexes from vertebrate cells and to begin characterizing the proteins. Human cells have provided the most detailed picture of the protein composition of these complexes. Considerable information is also becoming available for invertebrates, particularly *Drosophila*. Immunopurified 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 lower-abundance hnRNP proteins, and these appear to bind only to specific subsets of hnRNAs. Nuclease digestion experiments indicate that the hnRNP proteins are not simply dispersed on the hnRNAs, but rather are organized into interacting units.

The molecular cloning and sequencing of cDNAs for several RNP proteins led to the discovery of a conserved RNA-binding domain (RBD) and a ribonucleoprotein consensus sequence (RNP-CS). The RNP-CS has subsequently been found in RNP proteins from many sources. This octapeptide sequence, Lys/Arg-Gly-Phe/Tyr-Gly/Ala-Phe-Val-X-Phe/Tyr, is the most highly conserved segment in a generally conserved region of about 90–100 amino acids present at least once in many RNA-binding proteins of the nucleus and cytoplasm in organisms as evolutionarily divergent as yeast, plants, and humans. This region, because of its general conservation as a unit, was considered to be an RNA-binding domain, and the RNP-CS was suggested to be an important element in the RNA-binding activity of the RBD. These predictions have turned out to be correct, and the predictive value of consensus sequences for the identification of proteins as having RNA-binding activities has proved very useful.

We have produced the 94–amino acid RBD of the hnRNP C1 protein and purified it. The three-dimensional structure of this domain and of its complex with RNA oligonucleotide substrate are being determined.

Experiments on mitotic cells unexpectedly provided important insights into the assembly and general nature of hnRNP complexes. In mitosis, as the nuclear envelope breaks down, hnRNP proteins disperse throughout the cell, but they remain associated in complexes with RNA. After mitosis, once the nuclear envelope reforms, preexisting hnRNP proteins return to the nucleus. Double-label immunofluorescence microscopy with monoclonal antibodies to various

hnRNP proteins on postmitotic cells, however, revealed 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 snRNP proteins and lamins, are transported immediately (early group), while others, including A1, A2, B1, B2, E, G, H, and L, are transported later (late group). Thus immediately following reassembly of the nuclear envelope at the end of mitosis, pairs of cells are detected in which some hnRNP proteins are in the nucleus and others in the cytoplasm. These observations show that hnRNP complexes are dynamic structures in that hnRNP proteins can dissociate from the complexes and return to the nucleus separately.

It is possible that the association and dissociation of the complexes are controlled by post-translational modifications. An interesting candidate effector of such modifications is a cell cycle–regulated protein kinase that phosphorylates the hnRNP C proteins. The phosphorylation results in the appearance of novel, electrophoretically distinct forms of the proteins. These slower-migrating forms, which we term Cs proteins, are readily detectable by immunoblotting with anti–C protein monoclonal antibodies.

We partially purified the kinase (termed Cs kinase) that phosphorylates the C proteins to form the Cs proteins (several hundredfold) by high-performance liquid chromatography (HPLC). The activity of this novel cell cycle–regulated protein kinase is increased during mitosis. It is distinct from the histone H1 kinase component of maturation-promoting factor (MPF/*cdc*2). In addition to the Cs kinase, two other protein kinases were found to phosphorylate the C proteins. One of these is the previously identified casein kinase type II. The phosphorylated residues are serines, but the specific sites on the C proteins that are phosphorylated by the three kinases have not yet been determined.

Cs kinase activity was also detected in extracts of the yeast *Schizosaccheromyces pombe*. This should help in determining the relationship between Cs kinase and MPF/*cdc*2 and in understanding the function of the Cs kinase. It may also make it possible to identify the C proteins of *S. pombe*. Among the likely possible functions of the phosphorylations of the C proteins are a role in the assembly/disassembly of hnRNP complexes, an effect on the RNA-binding activity of these proteins, and a role in the transport of the C proteins to the nucleus.

Immunofluorescence microscopy with monoclonal antibodies to hnRNP proteins and photochemical crosslinking of RNA to proteins in living cells by ultraviolet light indicated that the A, B, C, L, M, and U proteins are nucleoplasmic—that is, they are nonchromatin proteins that are confined to the nucleus and excluded from nucleoli. The mRNP proteins that have been characterized so far are different from the known hnRNP proteins and are confined to the cytoplasm. Therefore the mRNAs must exchange most, if not all, of the proteins with which they are associated in the nucleus as they are transported to the cytoplasm.

The dissociation of these proteins from the nuclear mRNA and the subsequent binding of mRNP proteins must be an important aspect of nucleocytoplasmic transport. Understanding the mechanism that regulates the disassembly of hnRNP complexes prior to the transport of their proteins to the nucleus following mitosis may also be important for understanding this nucleocytoplasmic transport process. The most abundant protein that is bound to mRNA in the cytoplasm is the poly(A)-binding protein. We have isolated cDNA clones for it from several organisms, determined its sequence, and produced antibodies to it. We are currently studying its function.

Insulin Receptor Structure and Function

Leland Ellis, Ph.D.—Assistant Investigator

Dr. Ellis is also Assistant Professor of Biochemistry at the University of Texas Southwestern Medical Center at Dallas. He received his Ph.D. degree from the University of North Carolina at Chapel Hill with Aldo Rustioni and was a postdoctoral fellow at Columbia University College of Physicians and Surgeons with Karl Pfenninger and at the University of California, San Francisco, with William Rutter.

THE response of cells to the polypeptide hormone insulin begins with the binding of insulin to a specific cell surface receptor; this results in the rapid autophosphorylation of the cytoplasmic protein tyrosine kinase domain of the receptor specifically on several tyrosine residues. Although the postreceptor mechanisms involved in the mediation of the multitude of physiological responses to insulin remain largely obscure, this initial transmembrane signaling event is required for all known insulin-dependent cellular responses.

To begin to relate the structural features of the insulin receptor protein to the details of its function, we have established two experimental approaches. First, the availability of cDNAs that encode the insulin receptor protein, together with methods to introduce mutations into receptor coding sequences and to express these altered receptors in heterologous cell expression systems, has provided the tools with which to test ideas concerning the functional roles of particular receptor residues. Second, the size and membrane-associated nature of the protein have made it possible to exploit the relatively simple transmembrane topology of the receptor to study its two functional domains individually as soluble molecules: the extracellular domain is secreted as a heterotetramer that binds insulin with high affinity, and the cytoplasmic domain is an active protein tyrosine kinase.

The extracellular ligand-binding domain of the insulin receptor is a complex molecule: each half-receptor comprises 929 residues derived from both α- (735 amino acids) and β-subunits (194 amino acids) and includes 16 potential asparagine-linked glycosylation sites and 41 cysteines. There is little information at present as to how this domain folds during biosynthesis or how it interacts with insulin. The study of an extensive series of deletion mutants has revealed sites within the receptor primary sequence at which truncation results in the generation of independently folded soluble subdomains. These sites now provide landmarks that guide further biochemical and molecular dissections of this complex domain. Furthermore, the establishment of a heterologous cell expression system that provides tens of milligrams of a secreted soluble derivative of the extracellular domain provides sufficient protein for its further biochemical dissection and renders feasible efforts to crystallize this functional domain of the receptor, a prerequisite for the elucidation of its three-dimensional structure by x-ray crystallography, in collaboration with Wayne Hendrickson (HHMI, Columbia University College of Physicians and Surgeons).

The cytoplasmic protein tyrosine kinase domain of the receptor has been expressed independently of the cell membrane and as a soluble monomeric insulin-independent enzyme (401 amino acids). To be relevant to the study of the function of the kinase in the wild-type membrane-associated insulin receptor, the soluble kinase must exhibit functional properties characteristic of the wild-type receptor. In fact, this soluble enzyme is recognized by a panel of conformation-sensitive antireceptor monoclonal antibodies, it exhibits a low level of basal activity toward exogenous substrates that increases upon autophosphorylation, and the sites of autophosphorylation of the enzyme are typical of those observed following insulin-dependent autophosphorylation of the wild-type receptor in intact cells. Thus, although synthesized free of the cell membrane and now insulin-independent, this soluble derivative of the protein tyrosine kinase domain exhibits the two functional states observed for the kinase in the context of the wild-type transmembrane receptor.

An efficient heterologous cell expression system renders feasible the use of biophysical methods for functional studies of this soluble enzyme, which are not yet possible for the wild-type membrane-associated insulin receptor. For example, by the use of nuclear magnetic resonance (NMR) spectroscopy (in collaboration with Barry Levine, Oxford University), we have begun to explore the interaction of small molecules (metal ions, ATP, peptide substrates) with the enzyme and to study catalysis by the kinase in solution in

real time: we can now follow the phosphorylation of individual tyrosine residues of such peptides (especially the order of phosphorylation of multiple tyrosines), as well as examine the contribution of individual amino acid residues to the kinetics of tyrosine phosphorylation of the peptide. Thus it should be possible to determine directly the stereochemical requirements and dynamics of efficient exogenous substrate phosphorylation by this tyrosine kinase. These studies complement efforts to obtain the three-dimensional structure of the enzyme through x-ray crystallography (in collaboration with Wayne Hendrickson).

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 Associate Professor of Pathology and Biochemistry at the University of Oklahoma Health Sciences Center. 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.

LIFE in all large animals requires the capacity to ship oxygen and food to all organs and the cells within the organs. Blood serves this capacity, but the development of such a delivery system is not without its inherent problems. The blood vessels that carry the blood are subject to injury. When cut, bleeding is a complication. Fortunately, a complex system exists that allows the blood to clot at a wound site and prevent fatal bleeding in all but the most severe wounds. Relatively few people are troubled with an inability to clot blood. In contrast, the formation of unwanted blood clots either causes or contributes to nearly half of all deaths in developed countries. The major problems occur in heart attacks, strokes, pulmonary embolisms, and septic shock. The system that leads to blood clots involves more than 10 separate proteins and several different cells. To prevent unwanted clots, an equally complex collection of systems, referred to as anticoagulant complexes, has evolved to block or limit the blood clotting process. Our working hypothesis is that changes in the activity of these anticoagulant complexes may contribute to the formation of lethal blood clots.

Protein C, protein S, and thrombomodulin constitute one of the natural anticoagulant complexes that prevents unwanted blood clots. We have focused 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 the 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).

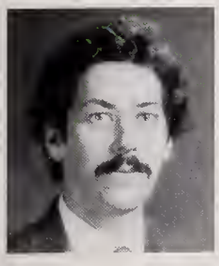
We found that only the free form of protein S could work to form the anticoagulant. Patients with clinical conditions known to cause an increased risk of blood clots also had reduced levels of free protein S and more C4bBP–protein S complex. Families with inherited thrombotic complications were identified 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 of throm-

bosis and examined whether elevation of the C4bBP levels to those found in patients would increase the risk of thrombotic complications (unwanted blood clots). This was exactly what was observed. More importantly from a therapeutic point of view, simultaneously increasing protein S prevented the thrombotic complications. These results suggest that if we could block the decrease in free protein S, the risk of thrombosis would be reduced. This approach could potentially return the patient to normal status without significantly increasing the risk of bleeding.

Recently it has become clear that clot formation in the arteries of the heart is responsible for most heart attacks. Damage that results from the attack may involve injury to the small vessels that feed the heart. When we examined whether the protein C system helps prevent this damage, we found that protein C is activated following a short (2 min or less) blockage of an artery in the heart. When this activation is prevented, the heart recovers less completely, if at all. These experiments imply that patients with decreased function of the protein C system may have increased damage to the heart if they do have a heart attack and that activated protein C may prove a useful agent in preventing some aspects of this damage to the heart. Clinical trials will test these concepts.

A major interest in our laboratory has been to understand how thrombomodulin causes thrombin to change from a clotting to a clot-inhibiting enzyme. To understand this question, we have made mutants of human thrombin, because the binding of thrombin to thrombomodulin probably changes thrombin's shape. This shape change probably accounts for how the complex accelerates protein C activation. One possibility is that thrombin has one or two specific amino acids that hit protein C in such a way that it cannot "dock" or bind with thrombin effectively. If this were the case, then altering the amino acids that block the docking event could make thrombin a better activator. We have examined two classes of such changes. In proteins, likes repel and unlikes attract. In the case of protein C, very close to the place where activation occurs, there is a negatively charged amino acid. There is also a negative amino acid in the region of thrombin very near where this negatively charged amino acid must dock. If the slow activation of protein C by thrombin is related to the charge repulsion (it is difficult to get negative charges together), then removing this negative charge from either thrombin or protein C should increase the rate of protein C activation in a manner analogous to formation of the thrombin-thrombomodulin complex. When either the negative charge on protein C or on thrombin is removed, protein C activation is increased 5–20 times. This change in thrombin does not influence the ability to clot fibrinogen. These studies provide insights into how regulatory proteins such as thrombomodulin may function and also provide the framework for the design, by selective mutation, of new anticoagulants that alter the function of normal clotting proteins.

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 Biomedical Sciences at the University of California, San Diego. He received his Ph.D. degree in microbiology and immunology from the University of California School of Medicine, Los Angeles. 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.

The Glucocorticoid Regulatory Network

The glucocorticoid receptor (GR) is well characterized as a potent activator of transcription. More recently the GR has also been shown to repress expression of a variety of genes in response to hormones. Glucocorticoids, in addition to their inductive effects on specific genes, inhibit proliferation of a variety of cultured cell lines and oppose the action of lymphokines in the activation of T cells. T cell activation is in part mediated by activation of Jun/AP-1, which is a member of the class of nuclear proteins encoded by proto-oncogenes and a factor in diverse aspects of cell growth, differentiation, and development.

We have provided evidence that the GR and Jun/AP-1 can reciprocally repress each other's transcriptional activation and that this inhibition is a consequence of protein-protein interaction. The interaction of two distinct regulatory proteins, referred to as "cross-coupling," has several important implications. First, it provides a molecular framework for understanding the opposing effects of growth factors and steroid hormones. Growth factors stimulate cellular division, whereas steroid hormones inhibit cell division while promoting differentiation.

Perhaps the most intriguing aspect of cross-coupling is that it allows interaction of distinct classes of transcription factors prior to their interaction with target DNA sequences. Thus the GR prevents AP-1 from binding to DNA and activating a downstream set of target genes. Accordingly, glucocorticoids indirectly regulate the AP-1 gene network by modulating the DNA-binding properties of Jun/AP-1. Although glucocorticoids are thought, classically, to regulate target genes by promoting the receptor to bind their DNA, cross-coupling allows steroid regulation of the AP-1-responsive genes, but in a DNA-independent fashion. This DNA-dependent pathway is a novel mechanism for steroid control of gene expression.

A Novel Retinoic Acid Response Pathway

The retinoids make up a group of compounds that include retinoic acid, retinol (vitamin A), and a series of natural and synthetic derivatives that exert profound effects on development and differentiation in a wide variety of systems. Retinoic acid has also been shown to induce the transcription of several genes, supporting the hypothesis that it has functions 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. At least three RXR-related genes (α, β, and γ) are now known to be located at genetically distinct loci. Each is capable of transactivating through cis-DNA-binding sequences similar to those of the RARs. 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. Binding experiments demonstrate that the RXR protein has low affinity for retinoic acid and, taken together with the transactivation studies, indicate that the RXR ligand may be a metabolite of retinoic acid. Thus a function of the RXRs may both overlap and be unique to that of the RARs. These studies suggest a role for RXRs in adult physiology and embryonic development.

The discovery in RXR of a second transduction pathway with distinct pharmacological properties may lead to a better understanding of how retinoids can elicit such an enormous diversity of biological responses and suggests that retinoid metabolism may provide important clues to the identification of new ligands.

In view of the homology of the RARs and RXRs, their genes would seem to have arisen independently during evolution, possibly to cope with the increasing demands of ever-more-complex evolving organisms. An understanding of the origins of these two gene families may afford insights to their modern roles. Thus, in search of invertebrate homologues, we have screened a *Drosophila* genomic library and identified the product of a novel steroid-receptor-like gene—*ultraspiracle* (*usp*)—that bears striking resemblance to RXR in both its DNA- and ligand-binding domains.

Genetic analysis indicates that the *usp* gene product has both maternal and zygotic functions. In the germline, it is important for formation of the embryonic cuticle choriogenesis and fertilization. In the embryo, it is necessary for larval pupation. Recent functional studies indicate that the *usp* gene product does not respond to retinoic acid, which suggests that a nonretinoid ligand may exist in *Drosophila*. This is particularly interesting in that its structure may provide clues to the identification of the vertebrate *usp* ligand.

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. A large number of agents act this way. They 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 actions of hormones and neurotransmitters are not due to changes in the cellular levels of cyclic nucleotide second messengers but result from an increase in the level of calcium ions. The next phase of the work involved the demonstration that these calcium-mediated agents act both by mobilizing calcium ions from internal stores and by stimulating the inflow of these ions across the cell membrane. We also demonstrated that the receptors for these agents are located on the outer surface of their target cells. Thus these cells must have some means of signaling from the receptors to their internal calcium stores.

Efforts were directed toward elucidating the signaling mechanisms. Initially we tested the hypothesis that the signal was generated by the breakdown of phosphatidylinositol, a phospholipid in the cell membrane. However, this breakdown was too slow to account for the changes in calcium. A related phospholipid, phosphatidylinositol 4,5-bisphosphate (PIP_2), broke down more rapidly in response to hormones, but the change was transient. The situation became clearer when inositol 1,4,5-trisphosphate (IP_3) was identified as the signaling molecule for calcium. 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 kinase (protein kinase C).

The present activities of the laboratory encompass three 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 protein from liver cell membranes and reconstituting it with other components of the signaling system. Reconstitution of the G protein with PIP_2 phospholipase C has been achieved, and the system has been used to purify the protein to homogeneity in both the complete form ($\alpha\beta\gamma$ heterotrimer) and in the form of the free α-subunit. The α-subunit (molecular weight 42,000) has been shown immunologically to be a member of the G_q family of G proteins.

The α-subunit activated by a GTP analogue has been used to identify the specific isozyme form of the PIP_2 phospholipase C controlled by the G protein. In both liver and brain, it is the 148-kDa β-isozyme. Evidence that the purified G protein is coupled to at least one receptor for a calcium-mobilizing agent has been obtained by the demonstration that binding of a GTP analogue to the G protein heterotrimer is enhanced by its reconstitution with the M_1 (calcium-mobilizing) muscarinic receptor for acetylcholine, but not with the M_2 form of this receptor. In addition, three calcium-mediated agents (vasopressin, epinephrine, and angiotensin) specifically stimulate the labeling of two G proteins by a radioactive, light-reactive analogue of GTP, azidoanilido GTP, in liver cell membranes. These proteins (apparent molecular weights 42,000 and 43,000) have also been shown immunologically to be members of the G_q family of G proteins. It is likely that the smaller protein is identical to the α-subunit (42,000 molecular weight) that activates PIP_2

phospholipase C. Possible functions of the larger protein are being examined.

The second major research area arose from the observation that hormones and neurotransmitters that break down PIP_2 in their target cells also break down another cell membrane phospholipid, phosphatidylcholine (PC). 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 in several ways: by mechanisms involving G proteins, Ca^{2+}, protein kinase C, and tyrosine kinases. *In vivo* and *in vitro* approaches are being used to explore these mechanisms.

The third major research area involves studies of the actions and metabolism of PA in cells. This work has shown that high concentrations of PA accumulate rapidly in cells stimulated with many hormones, neurotransmitters, and growth factors. Although it is widely believed that PA acts as a second messenger, its cellular targets and functions remain unknown. We have recently identified a protein kinase that is activated by PA and may therefore mediate its cellular effects. This enzyme is being purified and characterized.

The Molecular Genetics of Cancer

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.

OUR laboratory is studying the molecular basis of human cancer. In particular we are attempting to elucidate the earliest events that convert a normal cell to a malignant cell and the role of gene inactivation in cancer development.

The Earliest Events in Malignant Transformation

We have found that human cancer involves alterations of many genes, even within the same tumor. Most of these changes occur relatively late and probably do not play a causal role in initiation of malignancy. Cancer can require several decades to develop, including a long premalignant or benign phase. For example, in colorectal cancer, the time course of premalignant disease (adenomas or polyps) can be as long as 10 years. The only genetic change we have found in early adenomas is an alteration in DNA methylation—a tissue-specific modification of cytosine, one of the four DNA nucleotides. Hypomethylation, the loss of methyl groups from cytosine, appears to play an important role in normal gene activation or expression, and thus altered DNA methylation could contribute to the abnormal gene expression that characterizes cancer.

In addition to studying the problem of DNA methylation in cancer, we would like to know which genes are expressed specifically in cells that are transforming. By the time a cell is fully malignant, more than 1,000 genes have undergone changes in their expression. If we could determine which genes are activated in premalignant cells, we might better understand how structural changes in the genome mediate the dramatic changes in cell behavior that define malignancy, such as unregulated growth, tissue invasion, and metastasis.

To approach these two problems, we have developed a novel experimental system. We have captured cells in culture that have been treated with a carcinogen that causes hypomethylation of DNA. We have devised a way to tell which cells treated with this drug will become malignantly transformed, and which will not, before they have undergone any of the changes in growth properties or appearance that define malignancy. In this manner, we have isolated cells that look normal but are committed to malignant transformation, as well as cells that have been handled in the same manner but will not become malignant (the ideal control).

By preparing cDNA libraries (large collections of the expressed genes), we have identified a relatively small number that are specifically associated with commitment to neoplastic transformation. We have also identified several genes that are turned off early in that process. We are now determining the identity and function of these genes, one of which appears to be a novel oncogene with the ability to transform cells when introduced into them.

The Role of Gene Inactivation in Cancer Development

One of the most important areas of cancer genetics is the ascertainment and characterization of tumor-suppressor genes—genes whose *inactivation* contributes to cancer. Since almost all genes are present in two copies in the cell, inactivation of one copy could be transmitted in families from parent to child. Individuals inheriting one bad copy of the gene would be at increased risk of developing cancer. Cancer would develop from deletion or inactivation of both copies of these suppressor genes.

In the 1970s it was discovered that some children with Wilms' tumor, a childhood kidney cancer, are missing a large, microscopically visible portion of chromosome 11 (in band 11p13). Later we found that more subtle gene deletions on chromosome 11 could be detected in Wilms' tumors indirectly, using restriction fragment length polymorphisms (RFLPs), which can distinguish the maternal and paternal alleles, or copies, of a given gene (one on each chromosome).

During the past year, we have cloned 750 kilobases of DNA from band 11p13 of chromosome 11, using yeast artificial chromosomes (YACs), in collaboration with David Schlessinger and Maynard Olson (HHMI, Washington University) and

Bryan Williams (Hospital for Sick Children, Toronto). Within this chromosomal region, we found a cluster of genes that appear to be involved in both childhood kidney cancer and normal kidney development. In one-half of the tumors we examined, at least two of these candidate tumor-suppressor genes were turned off. Remarkably, the tumors in which the genes were turned off also contained tissues not normally found in the kidney, such as muscle, suggesting that the same genes are involved in both tumor development and normal kidney development. We now need to find out what role these genes play in kidney development and why they are turned off in cancer.

A surprising result from the cloning and mapping of this large genomic region was the identification of an "archipelago" of cytosine-guanine islands, or clusters of CG nucleotides that are often near genes. Several of the CG islands in the Wilms' tumor gene region were methylated, a feature normally characterizing the inactive X chromosome. This offers a possible mechanism for inactivation of tumor-suppressor genes, as well as an unexpected link to our research on DNA methylation.

In addition to the known gene on chromosome 11 that predisposes to Wilms' tumor, we have discovered another gene at a different location on chromosome 11 (band 11p15) that also predisposes to Wilms' tumor but is involved in bladder, breast, and lung cancer as well. Loss of one copy of this gene in the germline appears to cause Beckwith-Wiedemann syndrome, a disorder of organ overgrowth and malignancy. We are now attempting to clone this tumor-suppressor/Beckwith-Wiedemann syndrome gene. In this effort we are using a strategy we have developed to isolate tumor-suppressor genes directly. We fragment human chromosomes into subfragments of 2–10 million nucleotides in a way that allows us to transfer these into tumor cells. This novel technique will allow us to bridge a gap in cloning methods between chromosome-size pieces (average 100 million nucleotides) and yeast artificial chromosomes (average 300,000 nucleotides), and it may have general application to cloning a wide variety of genes.

Genetics, Structure, and Function of Histocompatibility Antigens

Kirsten Fischer Lindahl, Ph.D.—Investigator

Dr. Fischer Lindahl is also Associate 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 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 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 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, and almost every chromosome, including the mitochondrial genome, carries at least one.

The Maternally Transmitted Antigen

The sixth amino acid in the mouse mitochondrial protein ND1 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. Mta was the first minor H antigen to be analyzed in molecular detail. It has turned out to be an excellent model, although discovered by virtue of its unusual features of mitochondrial, hence maternal, inheritance and its presentation by a novel, highly conserved MHC molecule.

M3 only binds the ND1 peptide when the methionine at the end carries a formyl group, and M3 can also bind other peptides with a formyl-methionine. This is characteristic of the amino terminus of mitochondrial and bacterial proteins, and distinguishes them from proteins made in the cytoplasm of mammalian 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 have now cloned the gene *M3* from mice, and considerable effort has been spent during the past year on various systems for deriving M3 in amounts sufficient for a structural analysis. We can express the *M3* gene in insect cells, which make milligrams of the protein in a single culture flask. However, the protein is not expressed on the surface of these cells, probably because it is folded incorrectly, and is detectable only with rabbit antisera against short fragments of it. The heavy chain of MHC class I molecules is known to be unstable in the properly folded conformation at body temperature unless binding the β_2-microglobulin light chain as well as a peptide. We are now trying to achieve correct folding by mixing the M3 heavy chain from a cell extract with β_2-microglobulin light chains and Mta.

We can express *M3* in mouse fibroblasts, where Mta can be detected by killer T lymphocytes. The M3 protein is therefore folded correctly, but in amounts too small for biochemical analysis. This system does allow us, however, to change single amino acids in the protein by introducing mutations at specific sites in the gene. By testing whether killer T cells still recognize the mutant protein, we can assay whether it presents Mta. From 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 peptide.

RMA-S Mutant Cells

The RMA-S mutant cell line has been of fundamental importance in understanding peptide antigen presentation. The 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 synthetic peptides to the cells. It is thought to affect either an enzyme that cleaves proteins into peptides, or a protein that helps transport peptides from the cytoplasm into the endoplasmic reticulum, where MHC class I molecules fold.

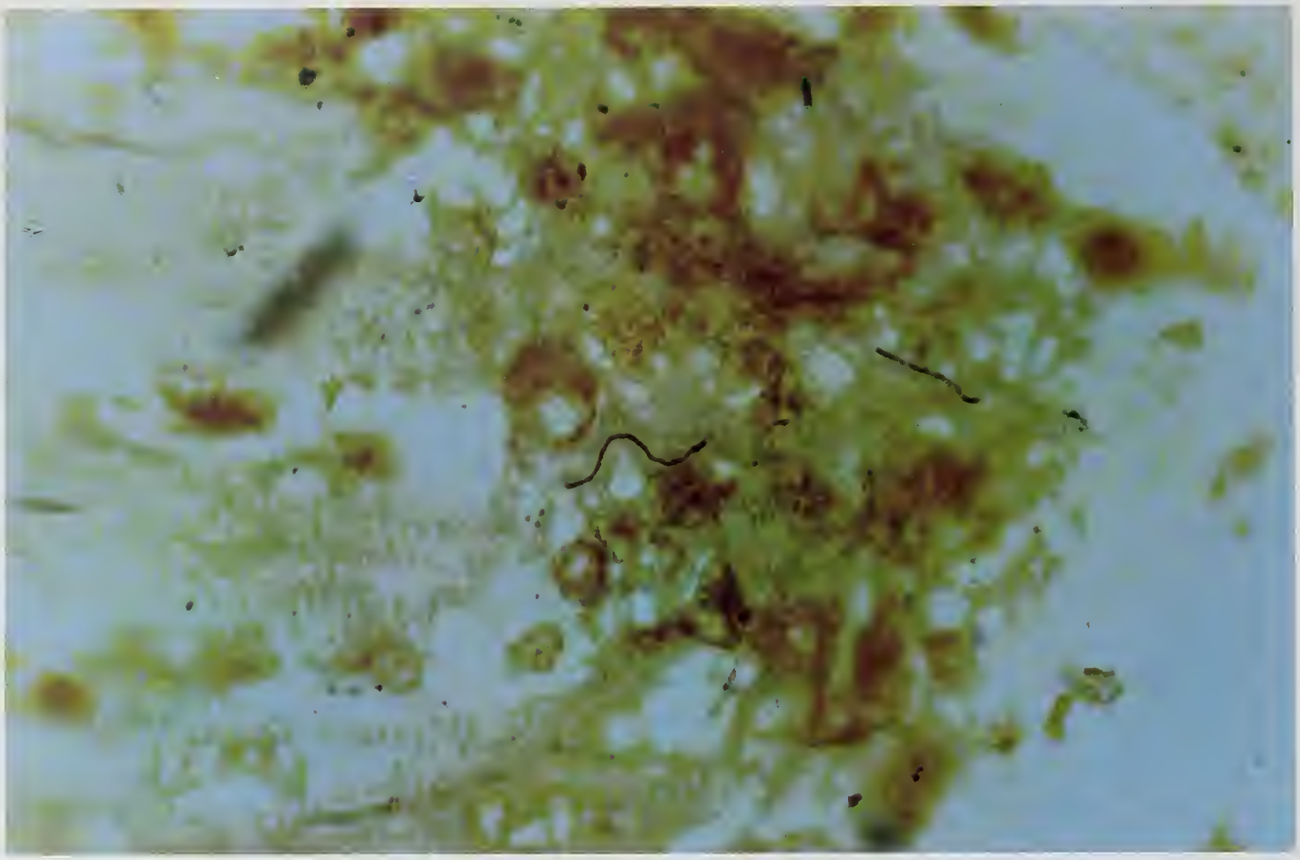
We have confirmed that these cells are devoid of classical MHC class I molecules that can be recognized by antibodies or killer T lymphocytes. But to our surprise, they still display detectable, albeit reduced, amounts of other MHC class I antigens, such as Mta and Qa-1. We infer that either these molecules bind their selected peptides with much higher affinity than the promiscuous classical MHC I molecules, or that these peptides are brought to the endoplasmic reticulum by a special pathway.

Most mitochondrial proteins are made in the cytoplasm, and much is known about how the mitochondria import them, but nothing is known about how proteins or peptides escape from the mitochondria. The Mta antigen is proof that mitochondrial peptides make it to the cell surface. We are now exploring mechanisms that may increase the production or release of mitochondrial proteins and peptides. We use both standard biochemical assays to test for the release of proteins from purified mitochondria and immunological assays on intact RMA-S cells to test specifically for the ND1 peptide. Antibodies against mitochondrial proteins are often present in diseases with an autoimmune component, such as primary cirrhosis of the liver. We hope to learn what brings about the initial exposure of these mitochondrial self-antigens to the immune system.

Mta in the Rat

To understand the importance and evolution of the Mta system, we look for parallels in other species. We have shown previously that the ND1 peptide of the rat can be recognized by mouse killer T lymphocytes if presented by mouse M3. We now know that the rat has its own maternally transmitted antigen system, which again involves a mitochondrial peptide. It differs in two important respects: the peptide is not from the amino terminus of ND1, and it is presented by the classical RT1A of the rat, not by a special MHC class I molecule.

The rat, however, does have an expressed MHC class I gene like the mouse *M3*. We have cloned and sequenced it. 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 observations 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.



Borrelia burgdorferi, *the Lyme disease agent, in heart tissue of infected mice. These rod-like bacteria are in the center and bottom of the picture and appear as a thin, dark line.*
Research of Richard A. Flavell.

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, where he worked with I. Gwyn Jones. His postdoctoral fellowships included research with Piet Borst at the University of Amsterdam and with Charles Weissmann at the University of Zurich. Before accepting his current position, Dr. Flavell was first Head of the Laboratory of Gene Structure and Expression at the National Institute for Medical Research, Mill Hill, London, and then President and Chief Scientific Officer of Biogen Research Corporation, Cambridge, Massachusetts. Dr. Flavell is a Fellow of the Royal Society and a member of several other distinguished societies.

MY laboratory has concerned itself for many years with the expression 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 approximately 25 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 soluble antigen that is 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.

Much of our work since joining HHMI has centered on the regulation of expression of the class II genes and the study of their biological role in transgenic animals. Class I and class II genes are both regulated *in vivo* by various lymphokines. 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.

In the past year we have taken a genetic approach to attempt to understand how interferon-γ activates the synthesis of these MHC molecules. A new approach has been used to isolate a series of mutant cell lines that are not capable of responding to interferon-γ. These cell lines appear to have a series of different defects, which we are attempting to elucidate. This genetic approach should help us understand how this important hormone regulates gene expression in this situation and elsewhere.

In studying the expression of class II genes, we have also analyzed class II protein function in an animal setting, employing the strategies of molecular genetics to perturb that expression and then observing the effect on immune function. A key issue in the functioning of the immune system is how the body discriminates its own tissues (self) from foreign components such as pathogens. An organism's inability to destroy self tissues is known as immune tolerance. Tolerance is generally believed to occur during the production of new T cells in the thymus by a process of clonal deletion; that is, self-reactive T cells are destroyed at the site of synthesis.

We have been interested in determining the mechanisms of tolerance to those components of the body that are never found 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, even though they appeared to be expressed specifically at these sites very distant from the thymus. Interestingly, it appears that 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 whether the above-mentioned mechanism applies to the more general situation of immune tolerance—namely tolerance to protein antigens themselves—we have generated transgenic mice that express the T cell receptor reactive with a specific antigen, the T antigen of SV40 (simian virus 40). We have obtained from several collaborators transgenic mice that express SV40 T antigen in various peripheral tissues in the body, including the β cells of the pancreas, the secretory tissues of the pancreas that produce digestive enzymes, and the lens of the eye. In the former two cases these experiments are analogous to our previous studies, except that now we are able to study directly tolerance to a protein antigen measured by its effects on a monoclonal population of T cells obtained by the use of T cell receptor transgenic mice. In the case of the lens of the eye, antigen is expressed in a compartment of the body traditionally believed to be an immunoprivileged site (inaccessible to the immune system). Tolerance in this compartment has not been previously studied. We have now bred the T cell receptor transgenic mice with those mice expressing antigens in these respective tissues and are studying tolerance mechanisms.

We have also initiated research to study the response of the immune system to the spirochete *Borrelia burgdorferi*, which causes Lyme disease, the notorious inflammatory disease. We have produced recombinant proteins for several of the outer-surface components of this spirochete and have immunized mice with these proteins. A mouse model has been developed by Stephen Barthold (Yale University) that exhibits many of the symptoms of the human disease, including arthritis and inflammation of the tissues surrounding the heart. Mice immunized with these proteins were challenged by infection with the Lyme disease agent. Strikingly, mice immunized with recombinant outer-surface protein A (OspA) were fully protected from infection with the spirochete. Both the protection and the immunity of the mouse to this organism appear to be long-lived. These encouraging results suggest that, administered appropriately, recombinant OspA could serve as a vaccine for protection of humans and animals against the Lyme disease agent. Vaccination may well be a desired approach, since this disease is widespread, causes serious chronic symptoms, and frequently goes undetected in early stages. We have also characterized the response of the human immune system to outer-surface proteins of this organism. The human immune response to OspA is qualitatively similar to the mouse response, suggesting that it should be possible to obtain a protective response in humans. Humans, however, respond to OspA later in infection than do mice, suggesting that further work may be necessary before an effective human vaccine is obtained.

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 Yale University School of Medicine. He received his B.S. degree in biochemistry from the University of Pittsburgh and M.Phil. and Ph.D. degrees in molecular biophysics and biochemistry from Yale University, in the area of x-ray crystallography. He carried out postdoctoral studies at Yale University in protein engineering and at Oxford University on the NMR spectroscopy of protein folding 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 working to understand the manner in which the amino acid sequence of a secondary structural element dictates its detailed conformation in the context of a folded protein molecule, using staphylococcal nuclease as a model protein system. We plan to test the generality of our observations and conclusions by examining the relationship between the amino acid sequence of immunoglobulin loops and their structure and the resulting ligand affinity.

We combine a number of methodologies to address these aspects of protein structure and folding, including x-ray crystallography, NMR spectroscopy, and molecular biology.

Mapping Structure in the Unfolded State of Proteins

Protein molecules in the unfolded and molten globular 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. Staphylococcal nuclease variants and fragments are used as a test system. 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 it can be used to generate hydroxyl radicals, which in turn cleave peptide bonds at positions in the protein chains in proximity to the chelator. The cleavage sites can be determined by peptide mapping and protein sequencing. Experiments designed to characterize the system and map distances in the unfolded chain are in progress.

Analysis of Protein Folding Using NMR Spectroscopy

In collaboration with Christopher Dobson's laboratory in Oxford, we have developed a series of NMR experiments to characterize the equilibrium-folding kinetics of staphylococcal nuclease. These methods have been used to investigate multiple folded and unfolded states of the protein. Applied to nuclease variants that differ in thermal stability, the methods have allowed us to discriminate between mutants that influence the protein folding pathway and those that modify the stability of the tertiary structure. By combining site-directed mutagenesis with NMR spectroscopy, we have been able to quantitate slow intramolecular equilibria, which should serve as a basis for the experimental quantitation of the forces that stabilize protein molecules.

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 understand better the detailed

structure of globular proteins and to define design principles for protein engineering.

We have developed a genetic system 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 the genetic data has led to a predictive model for this β-turn type in all globular proteins. This approach may be of general use in defining other sequence–secondary structure relationships.

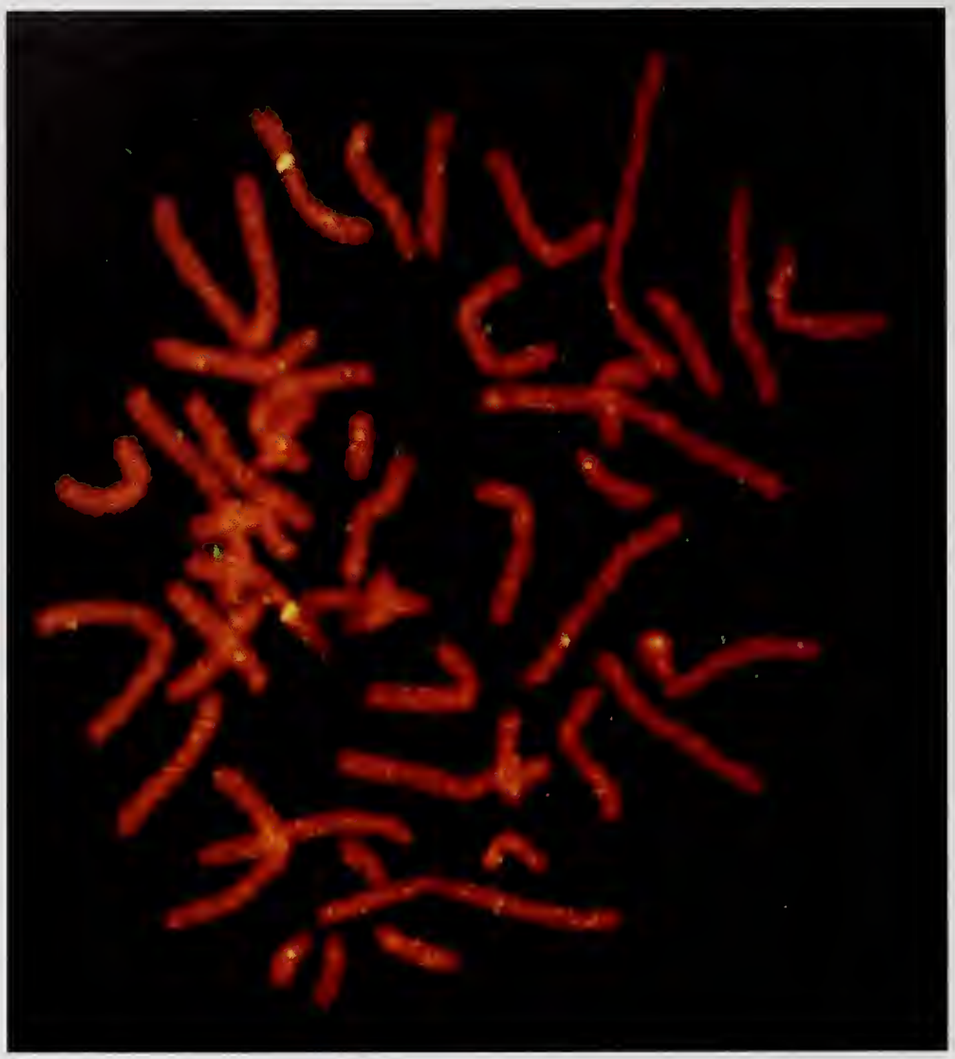
Crystal structure determinations of several mutants derived from a genetic analysis of this β-turn site indicate that substitutions can be made without greatly influencing the tertiary structure. In contrast, crystallographic analyses of several point mutants in another β-turn near the active site demonstrate that single–amino acid substitutions can result in a change from one β-turn type to another, causing a significant rearrangement in the local protein structure. We are currently working to understand the physical bases for the selections observed in the genetic experiment and the conformational alteration they impart to the protein. 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 Basis of Immunoglobulin Maturation

The sequence diversity found in immunoglobulin molecules is generated at several different levels. As B cells develop, combinatorial variability arises from the rearrangement of germline V, D, and J gene segments. When a B cell recognizes an antigen, somatic mutation of the variable region of the immunoglobulin genes is stimulated, adding further diversity to the immune response. Although these mechanisms have been well characterized, the structural basis by which sequence differences modify antibody affinities remains relatively unexplored.

Harden McConnell and his laboratory at Stanford University have prepared a panel of 12 monoclonal antibodies to a particular hapten (a small molecule that interacts with antibody but does not itself elicit an immune response) and have sequenced the heavy- and light-chain cDNAs. The antibody panel provides an opportunity to investigate the structural role of such rearrangements in determining antibody-hapten affinity.

We have crystallized a fragment of one of these monoclonal antibodies, with and without bound hapten, and have solved the structure of the complex, in collaboration with Axel Brünger (HHMI, Yale University). The crystal structure provides an opportunity to assess the effect of antigen binding on antibody structure and will serve as a basis for understanding the influence on hapten affinity of sequence variants resulting from gene rearrangement or somatic mutation.



Fluorescent in situ *hybridization to answer a clinical question. This shows chromosomes of a black man with tyrosinase-negative albinism. Since only one mutation in the tyrosinase gene was identified and the man was also mentally retarded, the question arose whether the other gene was lost through a submicroscopic deletion.* In situ *hybridization of a biotin-labeled cosmid clone containing the human tyrosinase gene revealed a signal (large yellow spots) on both chromosomes 11, where the tyrosinase genes reside, thus ruling out a deletion and forcing the conclusion that he is homozygous for the rare mutation.*

Cells and cosmid provided by Richard Spritz. Experiment done by Chih-Lin Hsieh, HHMI Associate at Stanford. Imaging on a Phoibos confocal laser scanning microscope.

From Spritz, R.A., Strunk, K.M., Hsieh, C.-L., Sekhon, G.S., and Francke, U. 1991. Am J Hum Genet *48:318–324.*

Human Molecular Genetics and Comparative Gene 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 physical and genetic maps of the human and mouse have undergone spectacular growth and development in recent years. Almost 500 homologous loci have been mapped in both species, and nearly 60 chromosome regions have been delineated that contain conserved syntenic groups of genes, through the efforts of many laboratories, including ours. Thus it has become possible, after mapping a gene in one species, to predict the location of the homologue in the other species. Comparative mapping information is also used to evaluate the possibility of a mouse mutation being a true model of a human genetic disorder.

Our laboratory is employing chromosomal *in situ* hybridization and 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 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—based on this knowledge—to devise precise diagnostic tests and rational treatment strategies.

Rett Syndrome

Rett syndrome is a neurodegenerative disorder that affects 1 in 15,000 girls. The affected girl is usually the only affected person in her family. Although she appears normal at birth and in her early infancy, her development slows and regresses in early childhood. The disorder is characterized by stereotypic hand movements, loss of speech, and profound mental retardation with seizures and microcephaly.

Although the progression of this disorder has been clinically delineated, the diagnosis remains tentative in the first few years. Pathophysiological and biochemical investigations have been unrevealing. The clinical picture is most consistent with progressive neuronal loss during early childhood, reaching a stable state that continues into adulthood. A genetic basis for this disorder has been established by concordance in identical twins and discordance in fraternal twins. A new mutation in a neuron-specific cis-acting gene on the X chromosome, which would be lethal in a male but allows females to survive due to their X chromosome inactivation mosaicism, appears to be the most plausible (although not the only possible) explanation of the underlying defect.

We are defining possible locations of the responsible gene by identifying the regions of the maternal X chromosome shared by affected half-sisters who have a common mother. To find candidate genes, we are carrying out chromosomal mapping studies of neuron-specific genes as they are cloned and become available. Genes that map to the X chromosome in a region not yet excluded from containing the Rett syndrome locus (such as genes encoding synapsin I, synaptophysin, and serotonin receptor lc) are being tested in 25 unrelated Rett families for structural rearrangements and sequence changes. Establishing the molecular defect in Rett syndrome will provide not only a much needed diagnostic test but also information about the role of the specific gene in the development and function of the human nervous system.

Inherited Dwarfism

Individuals with an autosomal recessive form of growth failure, despite having high levels of circulating growth hormone, are diagnosed as suffering from Laron dwarfism (LD) when they lack the activity of a specific growth hormone–binding protein in serum. They are also characterized by low levels of insulin-like growth factors that are normally produced by liver cells in response to growth hormone. Therefore the underlying defect in LD is thought to involve the growth hormone receptor (GHR). This receptor is encoded by a single gene, previously mapped by us to the proximal short arm of human chromosome 5 and to mouse chromosome 15. The sequence of the growth hormone–binding protein is identical to that of the extracellular domain of the GHR. Rearrangements of the GHR gene have been observed

only rarely in LD patients, which indicates molecular genetic heterogeneity of this disorder, and point mutations are yet to be identified.

Our laboratory is engaged in a collaborative study of a population in the mountains of southern Ecuador that is derived from early Spanish settlers and suffers from a high incidence of LD. Detailed clinical and biochemical studies of affected individuals by Arlan L. Rosenbloom (University of Florida, Gainesville) and his colleagues have revealed the classical features of LD, in addition to other manifestations not previously reported. We have established lymphoblastoid cell lines from several affected individuals and family members from different geographic locations. Because no gross rearrangements of the GHR gene have been identified, we have designed primers to use in the PCR (polymerase chain reaction) amplification of each of the 10 exons of this gene based on the published sequence. Single nucleotide changes in the amplified fragments are detected by denaturing gradient gel electrophoresis and confirmed by sequencing. A polymorphism in a GHR exon that cosegregates with the disease allele in this population has been detected and is currently used to develop a PCR-based allele-specific oligonucleotide hybridization test for the eventual detection of gene carriers. Our search for the molecular defect that renders the GHR gene nonfunctional is continuing. We are especially interested in learning whether a single mutation is responsible or whether more than one mutation contributes to the high incidence of LD in this population.

Another form of inherited dwarfism, autosomal recessive hypopituitarism, is due to a combined defect in three hormones produced by distinct cell types in the anterior pituitary gland: growth hormone, thyroid-stimulating hormone, and follicle-stimulating hormone. Besides failing to grow, these individuals fail to mature sexually and may suffer from hypothyroidism. We have begun a collaborative study of an inbred Hispanic family with several affected members. Our approach to unraveling the molecular defect is based on the human-mouse comparative map and on candidate genes. Two mutations in the mouse are phenotypically similar to this disorder: Ames dwarf (*df*) on mouse chromosome 11 and Snell dwarf (*dw*) on chromosome 16. By studying the inheritance of genetic markers on the human chromosome regions that are homologous to the relevant regions of mouse chromosomes 11 and 16, we hope to establish whether *df* or *dw* is a model for this disorder.

The candidate gene approach suggested that Pit-1, a pituitary transcription factor, may be a likely candidate. Indeed, recent studies from Michael Rosenfeld's laboratory (HHMI, University of California, San Diego) have documented a defect in the Pit-1 gene in two different mutant alleles at the *dw* locus. We have ruled out a structural rearrangement of the Pit-1 gene in our hypopituitarism family and have mapped the Pit-1 locus to a region of human chromosome 3. If the human disease gene cosegregates with markers in this region, we will analyze the Pit-1 gene for mutations. Alternatively, if the mutation is homologous to *df*, it should be linked to markers on either human chromosome 17 or 5q, and the responsible gene will be identified after it has been localized. This project exemplifies the goals of this laboratory—the generation of gene mapping information that is then used to identify and test candidate genes for human inherited disorders or mouse mutation models.

The Molecular Biology of Feeding Behavior

Jeffrey M. Friedman, M.D., Ph.D.—Assistant Investigator

Dr. Friedman is also Assistant Professor at the Rockefeller University. He received his B.S. and M.D. degrees upon completion of the Renssalaer Polytechnic Institute–Albany Medical College medical program. After completing a residency in internal medicine at Albany Medical College and a gastroenterology fellowship at Cornell University Medical College, Dr. Friedman enrolled in the graduate program at Rockefeller, where he received his Ph.D. degree for studies with James Darnell.

EXTENSIVE studies of humans and other organisms have suggested that body weight, body composition (percent body fat), and food intake are under strict physiologic control. The *set point hypothesis* predicts that both energy intake and expenditure are physiologically regulated in a particular individual to maintain a predetermined body weight. Implicit in this hypothesis is the notion that signal molecules that reflect the nutritional state of an individual are synthesized in the body (the periphery). The levels of these signals are sensed by control centers in the brain so that an appropriate response is generated to maintain a stable body weight. Studies in which separate regions of the rodent brain have been lesioned suggest that this feeding control center resides at least in part in the hypothalamus. The site of synthesis of the molecules that signal nutritional state is unknown, although fat cells, intestinal cells, or other cells of the gastrointestinal tract have been suggested. A knowledge of the site of synthesis and the molecular nature of the signals that affect brain centers controlling appetite could have important implications for our understanding of nutritional disorders in humans.

To understand more about these signaling mechanisms, we have been utilizing genetic and molecular biologic approaches to study the role of specific gene products in the control of feeding behavior and body composition. (The studies that aim to clone the mutant obesity genes *ob* and *db* from mice have been supported by the National Institute of Diabetes, Digestive, and Kidney Diseases.)

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 the animals' environment and to set up genetic crosses. In collaboration with Rudolph Leibel, we have begun a study of mice carrying recessive mutations that result in profound obesity. At least four different mutations in mice that cause obesity are available for study: *obese* (*ob*), *diabetes* (*db*), *fatty* (*fat*), and *tubby* (*tub*). In each case, a mouse becomes obese because of a defect in a single gene.

We have focused on the *ob* and *db* mutations for several reasons. Mutant *ob* and *db* mice become profoundly obese, often weighing three times as much as normal mice. As in humans, the obese phenotype in mutant animals appears to result from both increased food intake and diminished energy expenditure. Furthermore, Douglas Coleman at the Jackson Laboratory has suggested that *ob* mice are missing a circulating factor that suppresses appetite and that *db* mice, which are unable to respond to the *ob* factor, may be missing the receptor for this factor.

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 makes extensive use of restriction fragment length polymorphisms (RFLPs), which are genetic markers defined by specific cloned pieces of DNA. The first step in this approach makes use of genetic crosses between obese (or diabetic) mice and normal mice in which the pattern of inheritance of the obese phenotype is compared with that of RFLPs. Specific RFLPs that are inherited along with the obese phenotype are said to be linked genetically to the obesity gene.

By performing this analysis on several hundred mice with several dozen different RFLPs we have been able to identify a series of different RFLPs that are tightly linked to the *ob* and *db* genes. RFLPs that are genetically linked 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 have used three different RFLPs for several genes, including the *met* oncogene, a pancreatic enzyme carboxypeptide A, the *irp* gene, and the cystic fibrosis gene, to generate a detailed genetic map around the *ob* locus. Simi-

larly the mouse *db* gene has been mapped relative to RFLPs for interferon-α and a complement gene.

To identify other RFLPs that are more tightly linked than these probes, we have used the chromosomal microdissection technique: small slices of individual chromosomes are dissected and then cloned. Separate libraries have been made from proximal chromosome 6, where *ob* maps, and from mid chromosome 4, where *db* maps. Two probes from the chromosome 4 library that flank *db* are ~1 cM apart, and two probes that flank *ob* are ~1.5 cM apart. The proximity of these markers and their density will enable us to use evolving techniques in an attempt to clone the *ob* and *db* genes. Techniques such as pulsed-field gel electrophoresis and the cloning of large fragments in yeast artificial chromosomes are currently being used. The cloning of these genes should further our understanding of the mechanisms that control food intake and body weight.

Cholecystokinin Regulation, Function, and Expression 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 are important in the control of appetite.

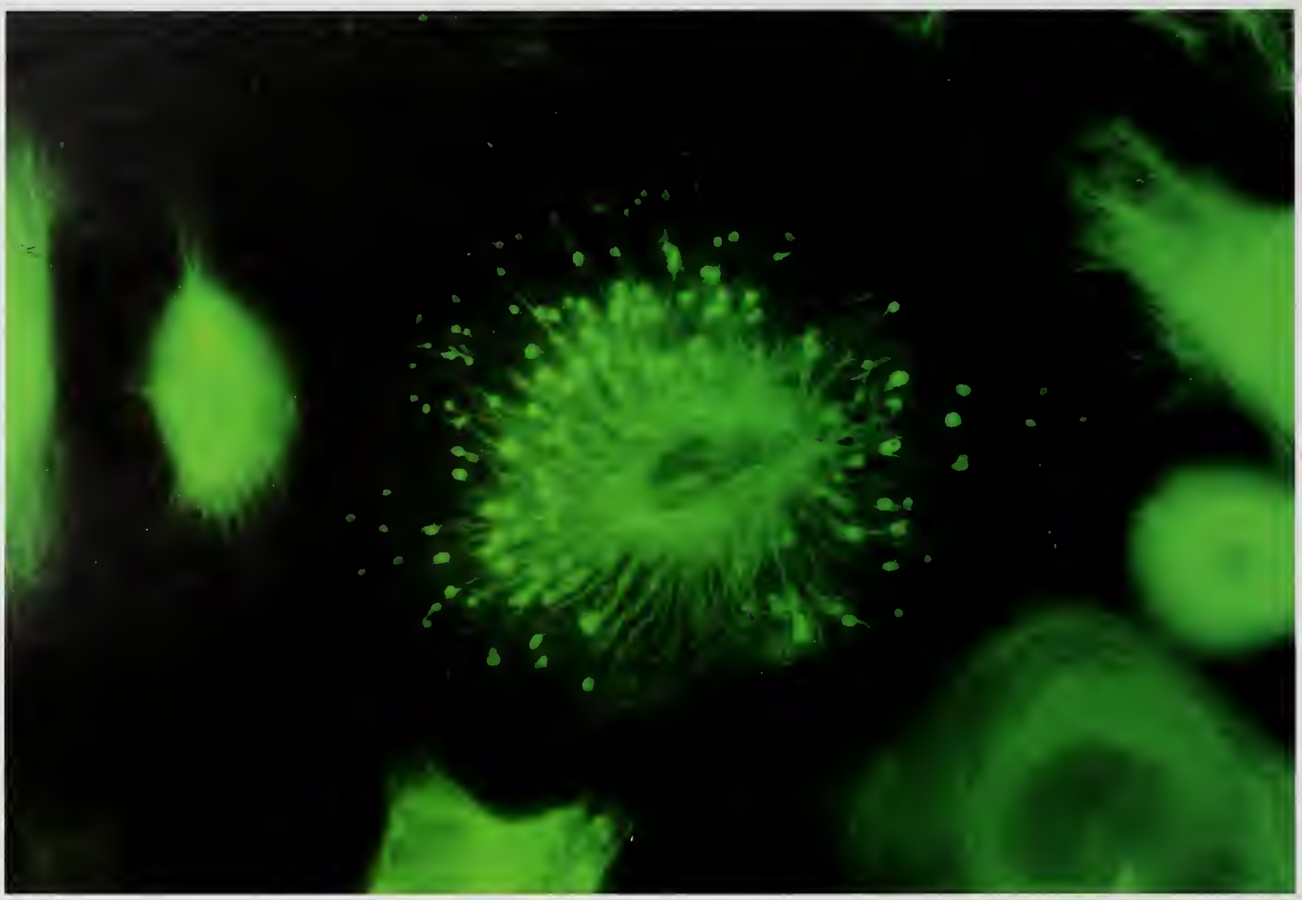
We have been conducting experiments to elucidate the molecular mechanisms controlling the expression of the CCK gene in both brain and intestine. In addition, we are using a variety of techniques to explore the function(s) of this hormone, with particular reference to the role of the CCK gene in the control of feeding behavior.

To define DNA elements involved in the regulation of the CCK gene, it was necessary to identify cultured cell lines that express this gene. Tumor cell lines derived from peripheral neuroepithelioma (a rare pediatric nerve tumor that usually develops in the chest wall) were found to synthesize CCK. Cell lines derived from another pediatric tumor, Ewing's sarcoma of bone, also express CCK mRNA, and subsets of other pediatric tumors, including rhabdomyosarcoma (a malignant muscle tumor), also appear to make CCK.

Characterization of these tumor cell lines also suggested that there is a class of pediatric tumors that express this hormone and that synthesis of this peptide may be of diagnostic and prognostic value in pediatric solid tumors. Of note, however, is that biologically active CCK is generated only after enzymatic cleavage of a precursor molecule (CCK prohormone) to an active form. The tumors that we have analyzed appear to synthesize the precursor but lack the ability to cleave it. The failure to note an association between these tumors and the expression of the CCK gene previously was likely a consequence of the failure of these tumors to synthesize processed CCK. Since the tumors do synthesize CCK mRNA and also secrete the CCK precursor, we have, together with Bruce Schneider, developed a novel radioimmunoassay that specifically detects the CCK precursor. In preliminary experiments it appears that the blood levels of the CCK precursor are elevated in patients with these tumors.

The identification of human tumor cell lines that express the CCK gene will make it possible to identify the DNA sequences required for cell-specific expression of this gene. These observations also suggest that CCK overproduction could in some cases have pathophysiologic effects in humans.

To describe the possible effects of ectopic CCK production on the control of feeding behavior, we have, in collaboration with Richard Palmiter (HHMI, University of Washington) and Ralph Brinster (University of Pennsylvania), artificially expressed high levels of this hormone in rodents by fusing the metallothionein (MT) promoter to the CCK-coding sequence and introducing the MT-CCK transgene into mice. Transgenic mice that express high levels of the CCK precursor in liver, plasma, and elsewhere are now available and have been characterized. Most of the tissues that expressed the CCK gene were similar to the human tumor cell lines, in that none of the expressing tissues processed the CCK precursor. Experiments to target CCK expression to tissues (such as stomach and pituitary) that can process the precursor to active forms are under way.



Cultured epidermal cells from a transgenic mouse expressing a mutant keratin gene. The cells were stained with an antibody specific to the mutant keratin protein. Note the presence of small balls, or aggregates, at the ends of the filaments. These structures are aberrancies in the filament network caused by the presence of the mutant keratin.

Research of Elaine V. Fuchs.

Regulation of Keratin Expression During Differentiation and Development in Human Skin

Elaine V. 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 my research is to understand the biochemical mechanisms that operate and regulate the expression of human genes during development and differentiation in skin. Our 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 probing the complex biochemistry of human skin. A major factor facilitating such studies is the ability to grow human skin cells—including epidermal cells, dermal fibroblasts, melanocytes, and dermal papillae cells—in tissue culture. These culture systems provide essential experimental models for studying many genetic skin diseases and skin cancers.

Much of our research on human skin has focused on the epidermis. The epidermis comprises about 20 cell layers, of which the 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 as yet unidentified trigger, 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 changes 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 keratin 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 closely 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 the keratins are ever expressed at any one time. As a normal epidermal cell differentiates, it changes the subset of keratins that 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. This new set of keratins is diagnostic for hyperproliferative abnormalities of the skin.

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 keratin mRNAs. We used DNA recombinant technology to show that the epidermal keratin mRNAs are encoded by about 20 different genes and that these genes are of two distinct types. Type I encodes small keratins (40–53 kDa); type II encodes larger keratins (53–67 kDa).

Keratins are expressed as specific pairs of type I and II proteins. In the past year, we showed that the basic subunit of keratin filaments is a heterodimer, composed of one molecule of each of the two keratin types. Approximately 20,000 heterodimers are necessary to form a single 10-nm filament; the assembly process is energy-independent and does not appear to require any auxiliary proteins or factors.

Using DNA sequencing, we have determined the amino acid sequences for several different keratin pairs. The sequences that are likely to be involved in the assembly process are highly conserved among keratin pairs, whereas the sequences that protrude along the surface of the keratin filament are different for different pairs of keratins. Hence the structures of keratin filaments assembled from different protein pairs are similar, but they are coated with different sequences and are likely to interact with different proteins inside the cell. In this way the cytoskele-

tal 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 molecular details of the filament assembly process and to investigate the nature of interactions of keratin filaments with other proteins and organelles, we have used deletion and site-directed mutagenesis to engineer alterations in the coding sequences of K5 and K14, the keratin pair expressed in the living cells of the epidermis. To generate substantial quantities of keratins for filament assembly studies, we used genetic engineering to overexpress wild-type and mutant human keratins in bacteria, which do not have keratin filament networks. We purified these keratins and examined the consequences of mutations and deletions on keratin filament assembly *in vitro*. We have begun to identify those sequences involved in filament elongation and those that are more important for lateral associations. In addition, we are using gene transfection of human tissue culture cells to examine 1) the dynamics of keratin filament assembly *in vivo* and 2) how the 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 into the function that these filaments perform in providing a relatively indestructible skin surface.

In addition to examining the structure of keratin filaments, we have also been interested in determining whether there might be natural mutations in human epidermal keratin genes that could lead to a genetic skin disease. We had previously shown that elevated epidermal keratin gene expression is a relatively late event in development and that certain keratin mutants have a dominant effect in cultured cells (i.e., they integrate into and disrupt the endogenous keratin filament network). In the past year, we made transgenic mice and used epidermal keratin promoters to target expression of some of these dominant mutant keratins to the mouse epidermis. Unexpectedly, the transgenic mice exhibited a phenotype that strongly resembles that of patients with epidermolysis bullosa simplex (EBS), a dominant and sometimes devastating, blistering human skin disease of unknown etiology. We are currently examining patients with EBS to determine whether they have mutations in their epidermal keratin genes.

We are also analyzing the complexity of the multiple keratin genes in the human genome and the mechanisms underlying their differential regulation. We have already shown that keratin expression is controlled at the transcriptional level and that it is regulated by vitamin A. The two genes expressed in the basal layer of the epidermis have been isolated and characterized, and we are focusing on how these genes are regulated during development and how their expression is influenced by retinoids. In the past year, we identified proximal and distal domains that synergistically act to regulate expression of one of these basal genes. Our goal is to identify the sequences and transcription factors involved in the regulation of these genes. This analysis of the factors controlling the cell's major structural genes should lead us to the factors involved in determining keratinocyte fate. Unraveling the nature of promoter and enhancer sequences involved in regulating epidermal genes will not only be important for understanding epidermal development but may also be useful for targeting products to the epidermis, e.g., in drug therapy. Because epidermal cells can be removed from a patient, cultured *in vitro*, and grafted back onto the patient, such techniques are potentially useful for medical research.

Knowledge of the normal regulation processes of epidermal and hair differentiation will be essential to identify the points at which these processes go awry in different genetic skin diseases and skin cancers. Elucidating the molecular nature of the normal and abnormal programs of differentiation in the skin should lead to new and badly needed treatments for dermatological diseases.

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 peptide-releasing 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. It was 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 the sperm cell 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. The small molecule cGMP 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. Although this receptor also will bind ANP, high concentrations of the peptide appear required for enzyme activation, which suggests the existence of a more potent natural peptide.

We also have recently 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 ANP 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.

These results suggest that drugs specific to different forms of guanylyl cyclase may result in specific physiological effects. 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.

Detection of Genetic Recombination in Germ Cells

Jan Geliebter, Ph.D.—Assistant Investigator

Dr. Geliebter is also Assistant Professor and University Fellow 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. 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 research 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 testing thousands of mice with skin grafts. 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 a novel approach to detect microrecombinant *H-2* genes. Mice with microrecombinant *H-2* genes arise from germ cells (sperm or egg) that contain microrecombinant *H-2* genes formed in the gonads (ovaries or testes) of normal mice. Therefore, if one looks in the sperm or eggs of normal mice, one should find microrecombinant *H-2* genes. We are using this approach to investigate the microrecombination process. Since we can obtain thousands of eggs or millions of sperm from each mouse, we can analyze the equivalent of millions of mice for microrecombinant *H-2* genes. We are using the polymerase chain reaction (PCR) to amplify specific *H-2* genes and then clone the genes. Small DNA probes are being utilized to identify the clones containing microrecombinant *H-2* genes.

We are presently analyzing the microrecombination frequency in ovarian cells, as data from previous studies indicate that microrecombinations occur in female germ cells. We also hope to determine whether microrecombinations occur in sperm cells as well, and at what frequency. (Evidence from skin graft studies suggested that microrecombination may occur only in female mice.) Some strains of mice may undergo micro-

recombinations more frequently than others. Our goal is to determine microrecombination frequencies in several mouse strains and perhaps identify critical parameters in the microrecombination 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 *H-2* genes, *Qa* and *Tla* region genes are characterized by sequence conservation among alleles and limited polymorphism. The lack of polymorphism has been suggested to preclude an immunological function for the products of these genes. We have identified a *Qa* region gene whose sequence differs greatly between alleles of the C57BL/6 and C3H mice. The sequence differences between the two alleles is manifested in both scattered and clustered nucleotide substitutions. The clustered substitutions, which are similar to those observed in the microrecombinations that diversify *H-2* genes, 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.

PCR analysis has indicated that this *Qa* gene is transcribed in some strains of mice. We are engaged in an in-depth analysis of the transcription, translation, and cell surface expression of this gene and its product in order to ascertain its function. This gene is polymorphic in at least three strains of mice, and other strains are being analyzed. 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.

Genetic Analysis of Growth Control

Donna L. George, Ph.D.—Assistant Investigator

Dr. George is also Associate Professor of Human Genetics at the University of Pennsylvania School of Medicine. She received her Ph.D. degree in zoology and genetics from Rutgers University. Her postdoctoral research was done at the Oak Ridge National Laboratory, where she characterized basic mechanisms associated with the ultraviolet induction of mutagenesis in Escherichia coli. *After completing a postdoctoral research program with Uta Francke at the University of California, San Diego, she was Assistant Professor in the Department of Medicine at the Johns Hopkins University School of Medicine, where she investigated chromosomal aberrations and their role in tumorigenesis.*

CELLULAR growth is an intricately regulated process that is tightly controlled by the coordinated interactions of many genes. Mutations that alter the structure or expression of a broadly based class of genes can result in the perturbation of normal cell proliferation; such uncontrolled cellular division is a hallmark of cancer. To understand better the molecular events that dictate the complex course of tumor development and progression, we have focused on the identification of these critical growth control genes and characterization of the mechanisms by which they are regulated.

Oncogenic Potential of a Novel Amplified Gene

Studies performed in this and other laboratories have provided substantial evidence that the physical amplification and consequent overexpression of some growth control genes are associated with the initiation and/or progression of a variety of mammalian tumor cells. A functional analysis of amplified DNA in tumor cells should allow the identification of other genes that have a central role in the regulation of normal cellular proliferation and that may contribute to cellular transformation. We have identified a candidate for such a gene. Amplified DNA sequences present in the tumorigenic mouse cell line 3T3-DM have been analyzed to determine whether the presence of cellular transforming activity is correlated with the elevated expression of any of the amplified genes. Our results have provided evidence that the *mdm-2* gene, which is amplified more than 50-fold in the 3T3-DM cell line, induces tumorigenicity when overexpressed in established, nontumorigenic cell lines. Analysis of the predicted amino acid composition of the *mdm-2* product reveals features that are similar to those that are functionally significant in certain DNA-binding proteins. Thus *mdm-2* may participate in growth control pathways by affecting the expression of other genes. We are raising antisera directed against the product of this evolutionarily conserved gene to facilitate analysis of its normal cellular function.

Expression of the c-*Ki-ras* Proto-oncogene

Despite its potentially pivotal role in growth control pathways, relatively little is known about the genetic elements that mediate the expression of the c-*Ki-ras* proto-oncogene. Because this gene is normally expressed constitutively at low levels in all cell types, it is generally considered a member of a class of "housekeeping" or growth control genes. Using the c-*Ki-ras* gene as a prototype, we have carried out experiments to identify mediators of expression that might be common to these genes. Functionally important domains within the transcriptional control (or promoter) region have been identified, and we have begun to examine specific DNA sequence motifs that may represent binding sites for particular nuclear regulatory factors.

One promoter element exhibits a particular transcriptional influence. This element has an unusual sequence composition (rich in C + T nucleotides on one strand, referred to as a pur/pyr domain) and binds one or more nuclear factors. Several lines of evidence indicate that the same or related DNA-binding proteins recognize similar pur/pyr domains that are located in the promoter regions of other growth control genes. Thus the sequence similarities of these regions may reflect a functional relatedness. The potential of the pur/pyr regions to adopt unusual structural conformations may also have implications for the expression of genes in which they are contained; such elements may allow for the maintenance of these promoters in a structure poised for transcription.

We have isolated recombinant DNA clones representing genes encoding DNA-binding proteins that recognize the pur/pyr motif. The availability of these clones will allow the *in vitro* and *in vivo* studies needed to define the role that these unusual sequence elements may play in mediating gene expression control.

The Genesis of Meningiomas

A new avenue of investigation is aimed at elucidating the molecular changes involved in the de-

velopment of meningiomas, a specific class of brain tumors that originate in the thin membrane covering the brain and spinal cord. Meningiomas are one of the most common tumors of the central nervous system, accounting for up to 20 percent of all intracranial tumors and 25 percent of intraspinal neoplastic lesions. Currently the mechanisms responsible for the initiation and progression of these tumors are largely unknown.

Extensive cytogenetic analyses, coupled with recent DNA studies, suggest that meningiomas result from the loss of function of a tumor suppressor gene, the meningioma susceptibility locus (*msl*), located on human chromosome 22. It is proposed that both alleles of this gene must be lost or functionally inactivated for tumorigenesis to occur. Although only a few such genes have been isolated, it has become clear that the protein products of tumor suppressor genes play critical roles in negatively regulating cellular proliferation. We have initiated studies to identify the *msl* gene and other functionally important genes whose expression is missing or altered in meningiomas relative to their normal precursor cells.

To do this, we have employed subtractive hybridization and cloning. These protocols remove sequences that are common to two cell populations and enrich for sequences that are present in only one of these populations. Utilizing cultures of primary leptomeningeal cells established in this laboratory and two meningioma cell lines, we have constructed subtraction cDNA libraries whose cDNA contents are enriched for sequences of genes that are expressed in normal human leptomeningeal cells but are either not expressed or aberrantly expressed in meningioma tumors. Screening of these libraries has identified, and will continue to provide, candidate clones for the *msl* tumor suppressor gene and other genes of interest. More detailed characterization of these clones is in progress.

The isolation of the *msl* gene and downstream affected genes has diagnostic and therapeutic implications for many tumors of neural crest origin. The availability of the leptomeningeal cell lines and genetic markers will provide a unique set of tools to begin dissection of the critical genetic events associated with the etiology and pathology of meningiomas and the elucidation of what is expected to be a fundamental pathway in the development and differentiation of multiple cell types in the central nervous system.

DNA-sequencing Technologies

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. He was an NIH postdoctoral fellow at the Institute de Biologie Moléculaire, Geneva, Switzerland. Dr. Gesteland served as Assistant Director and Investigator at Cold Spring Harbor Laboratory, New York, before assuming his present responsibilities.

THE elegant and now widely used methods for sequencing DNA are precise and quite simple, but labor-intensive, expensive, and frustrating for large projects. The current automated instruments based on fluorescence detection methods are clearly a step forward, but another enhancement of efficiency is needed to bring large projects like sequencing of the human genome within range. The heart of sequencing is the remarkably powerful size-fractionation of DNA chains into ladders with one-base resolution. In trying to evolve more efficient, ladder-based methods, we have pursued two approaches.

Many of the labor-intensive steps of sequencing, from growth of bacterial cultures and DNA preparation to fractionation of samples on the crucial gels, can be made parallel processes by multiplexing, as pioneered by George Church (HHMI, Harvard Medical School) using chemical sequencing and by us using chain termination sequencing. Here, many vectors, each with its insert of DNA to be sequenced, are combined at an early stage and sequenced and fractionated as a mixture. Each vector in the mix has the same sequence to prime DNA synthesis into its insert, so that all are done in one simple reaction, but each has unique sequences as tags just flanking the insert.

The ladders of mixed sequences fractionated on a gel are transferred to membranes, which are then probed sequentially with radioactive or chemiluminescent probes that recognize each of the tag sequences, revealing one sequence pattern after another. The labor-intensive steps are simplified by the level of multiplexing (10- to 40-fold), and the readout is reduced to repetitive rounds of probing that can be readily automated. Various architectures for automation and signal detection are under development.

The bottleneck for multiplex sequencing now shifts to other fronts. The "back end" problem of reading the DNA sequence off of the autoradiograms that result from each daily cycle of probing quickly becomes rate-limiting. Automation of this crucial sequence-calling is difficult and has been getting much attention. We are trying a new approach based on communication and signal-processing theory.

The "front end" problem of feeding appropriate sequences into the system is especially crucial. In the past, multiplex sequencing has been done with random fragments of DNA in the hope of reassembling the large sequence by overlapping information from 5- to 10-fold redundant sequencing. This becomes increasingly difficult with increasing project size, as the gaps are difficult to fill. An organized front end seems necessary. Thus we are developing *in vivo* approaches to generate and map the input inserts.

Clones to be sequenced are generated in bacterial cells by turning loose a transposon that inserts itself quite randomly into the large DNA of interest. Quick-mapping methods allow assembly of the minimally overlapping set of clones to ensure coverage. The transposons are designed to have all the crucial elements for multiplex sequencing in both directions out from the point of insert, thus generating twice the reading length. Small-scale tests are in progress and should tell us if this front end approach will have the expected simplification for multiplex sequencing.

A second approach to sequencing that might provide increases in efficiency and speed is capillary gel electrophoresis. Recent developments here and in other laboratories have shown that microbore capillaries with an internal diameter of 50–70 microns can be used to fractionate DNA samples by size, just like the conventional slab gels. The virtue of capillaries is that very high voltages (10–30 kilovolts) can be used, since heat dissipation is so efficient. Molecules 300–400 in length can be analyzed in minutes, with a resulting increase in resolution.

The difficulty is that the small gels can accept only small amounts of DNA, so that detection must be very sensitive. Working with Norm Dovichi (University of Alberta, Edmonton), we have adapted a sheath flow cuvette for detection of four fluorescent tags so that detection in the

range of a few thousand molecules is achieved and all four bases can be detected in the output from one column. Charge-coupled devices (CCDs) are ideal for detection of these low-fluorescent signals and permit spatial resolution that will be helpful for continuous spectral analysis and for parallel analysis of multiple capillaries.

Commercial instruments for conventional zone capillary electrophoresis have worked out automated sample-handling that can be adapted to the sequencing problem. We see these automated and multiplexed capillary sequencing instruments as potential alternatives to current fluorescent machines for primer-based sequencing.

We have been struck by the interest that colleagues in more engineering-related sciences have shown in these problems. The project has resulted in a refreshing blend of interests that will bring new solutions to molecular investigations.

Membrane and Secretory Proteins

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 at 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, for three years before going to the University of Texas.

INVESTIGATIONS in this laboratory focus on the molecular genetics of membrane and secretory proteins. Experiments involve three proteins: the hemagglutinin (HA) of influenza virus, which serves as a marker molecule for specific cell populations in transgenic mouse experiments; human tissue-type plasminogen activator (t-PA), a serine protease produced in vascular endothelial cells and involved in thrombolysis (dissolving blood clots); and BiP/GRP78, a protein of the endoplasmic reticulum (ER) that is involved in the initial mobilization of proteins that traverse the secretory pathway.

Development of a Model for Autoimmune Type I Diabetes

In type I diabetes the destruction of insulin-producing pancreatic β cells is thought to occur via an autoimmune mechanism—an immune response directed against the host. Surface molecules of the β cells have been implicated as target antigens. We have used HA as a marker molecule to investigate the mechanism of immunological tolerance and to develop an animal model for the study of diabetes.

In collaboration with Robert Hammer (HHMI, University of Texas Southwestern Medical Center at Dallas), we have microinjected the HA gene—placed under the control of sequences from the rat insulin II gene—into fertilized mouse embryos to generate three lines of transgenic RIPHA mice in which the transgene is expressed only in the β cells of the pancreas. We have now bred large numbers of transgenic mice that are homozygous for the HA gene and for the $H\text{-}2^d$ major histocompatibility gene.

From birth these transgenic animals have slightly raised blood glucose levels compared with those in control animals. Although histological analysis of the pancreata of young RIPHA mice reveals some disorganization of the normally ordered architecture of the islets, these animals display no physiological problems until approximately 4–5 months of age. However, after this time up to 27 percent of the RIPHA mice develop frank diabetes mellitus as a consequence of lymphocyte-mediated autoimmune destruction of the pancreatic β cells. Analyses of the sera of diabetic animals often reveal the presence of antibodies directed against HA, as well as antibodies that recognize proteins of the β cells. The high blood glucose levels in the diabetic mice can be decreased to normal levels by administration of insulin. Therefore the development of hyperglycemia in these transgenic mice has all the features that typify the pathogenesis of human type I diabetes mellitus. These mice should provide a useful animal model for further studies of this disease.

The development in RIPHA mice of immune responses against HA raises the question of why the protein, which is expressed on the β cells of the developing pancreas before thymic maturation, is not recognized as a self antigen. Additional lines of transgenic mice have now been developed that express HA under the control of the mouse metallothionein promoter. These MTHA mice, which express high levels of HA in many tissues (including liver and kidney), will enable investigation of the immunological responses to HA when the protein is expressed in major cell populations. Thus far, MTHA mice up to 10 months of age have displayed no signs of diabetes or any other autoimmune disease.

Human Tissue-Type Plasminogen Activator

Many normal and abnormal processes requiring extracellular proteolysis are thought to be mediated by plasminogen activators that cleave plasminogen to the active protease plasmin. Evidence points to t-PA as the physiological thrombolytic agent in the vascular system.

The level of t-PA activity in the circulation is controlled by the interaction of the molecule with three other proteins. First, the affinity of the enzyme for its substrate plasminogen is increased several hundredfold by binding to fibrin. Second, t-PA is rapidly inactivated by the serpin, plasminogen activator inhibitor-1 (PAI-1). PAI-1 acts as a suicide substrate and forms a covalent bond with a serine residue in the active site of t-PA. Finally, the enzyme is efficiently cleared from the blood

circulation by specific t-PA receptors on liver cells.

The t-PA polypeptide is composed of a number of independent functional domains, which include a hydrophobic signal sequence, a short pro segment, a finger domain having homology to the fibrin-binding finger domains of fibronectin, an epidermal growth factor (EGF)-like domain, two kringle structures, and finally the catalytic domain that is homologous to other members of the serine protease family. We have shown that the finger and/or EGF-like domains are involved in the initial high-affinity binding of t-PA to fibrin, while stimulation of t-PA activity involves secondary, lower affinity interactions of fibrin with either of the two kringle domains. The binding of t-PA to specific receptors on hepatic cells also involves sequences within the finger and/or EGF-like domains.

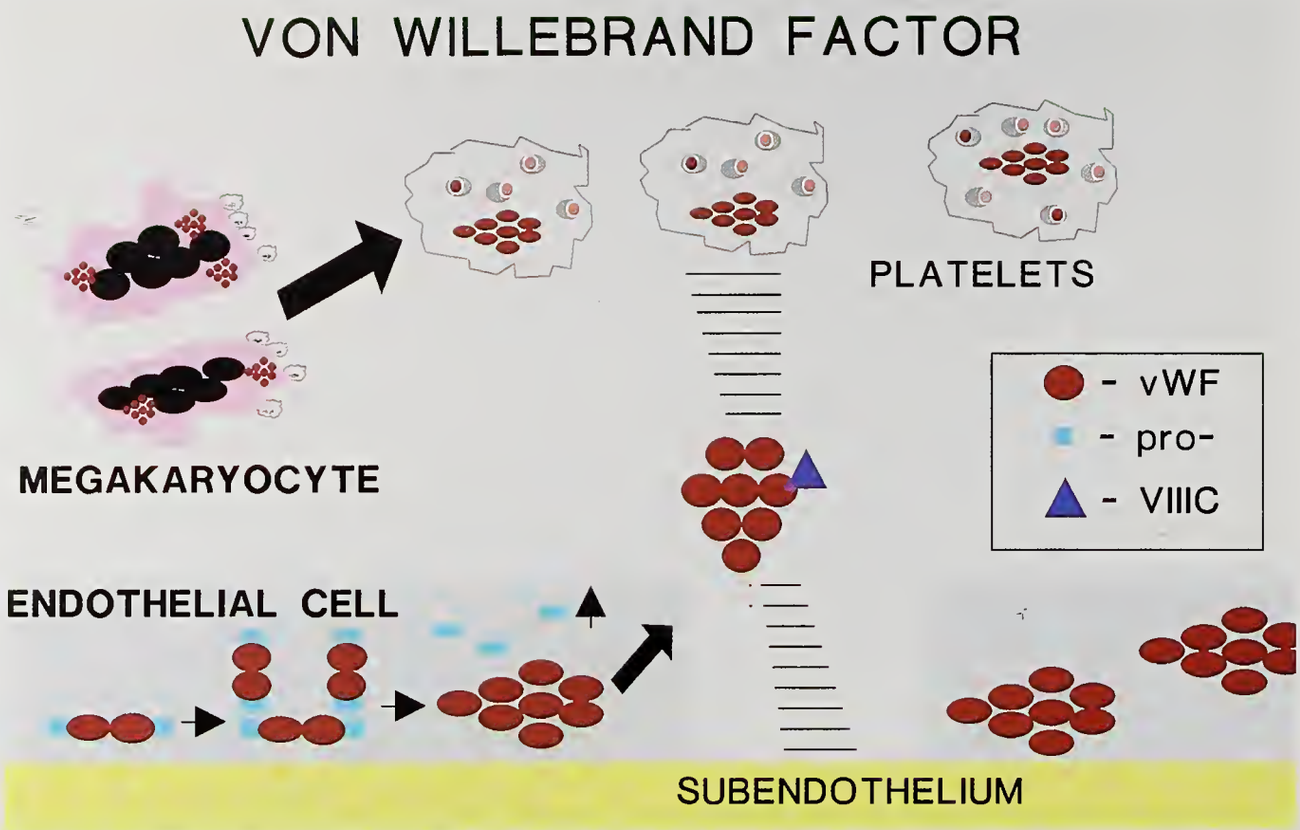
Although the three-dimensional structure of t-PA has not been elucidated, we have been able to model the finger domain, the EGF-like domain, the kringle domains, and the light-chain/inhibitor complex using the known structures of homologous proteins. Site-directed mutants designed using these proposed structures have provided information about the individual amino acid sequences that interact with the effector molecules. For example, we have generated variant enzymes 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.

Protein Folding Within Cells

Until recently it was widely assumed that the folding and assembly of newly translocated polypeptides into their tertiary and quaternary structures is a spontaneous process that does not involve the intervention of other cellular proteins. It is now apparent, however, that members of heat-shock protein families, including the ER-resident protein BiP, are intimately involved in facilitating the folding and assembly of nascent polypeptides.

BiP binds transiently to a variety of wild-type membrane and secretory proteins and more permanently to malfolded proteins that are trapped in the ER. These observations led to the hypothesis that BiP plays a role in folding in the ER lumen. Analysis of the interaction of BiP with HA mutants that lack various structural domains has demonstrated that BiP binds to amino acids that form the trimeric stalk domain in the folded HA molecule. BiP also appears to bind to and stabilize partially folded polypeptides in a state competent for further folding and oligomeric assembly.

Cloning and sequencing of cDNAs encoding BiP from mammalian and yeast cells have revealed an extremely high degree of evolutionary conservation. The yeast BiP gene is essential for cell viability and, like its mammalian counterpart, can be induced by the accumulation of unfolded proteins in the ER. The yeast gene differs from the mammalian gene, however, in also responding to heat shock. All the information required for accurate transcriptional regulation of yeast BiP is contained within a 230–base pair region directly upstream of the initiation codon. This sequence contains a functional consensus heat-shock element and an element that is responsible for induction of yeast BiP mRNA following accumulation of unfolded proteins in the ER. This unfolded protein response (UPR) element displays homology to conserved sequences present in the promoters of mammalian glucose-regulated proteins.



Schematic illustration of the biosynthesis and function of von Willebrand factor. Red ovals represent the factor, which is synthesized in the bone marrow megakaryocyte (precursor of blood platelets) and in the vascular endothelial cell. It is assembled through several steps into a large complex containing up to 100 subunits. Bound to von Willebrand factor, in the circulating blood, is the antihemophilic factor VIII (blue triangle). In addition to carrying factor VIII, von Willebrand factor serves as a major adhesive link between platelets and the blood vessel wall at sites of injury.

Research of David Ginsburg.

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 the Brigham and Women's Hospital, Harvard Medical School.

THE major research activities of my laboratory are centered around the study of 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, which is an important part of the body's blood clotting system, is to serve as a bridge connecting blood platelets to the wall of injured blood vessels, thereby helping to control bleeding. von Willebrand factor also serves as the carrier for factor VIII, the substance missing in patients with hemophilia. Abnormalities in von Willebrand factor result in von Willebrand disease, the most common human inherited bleeding disorder, occurring in 1–3 percent of the general population. Although more than 20 different subtypes of this disease have been described, the molecular basis for this variation is unclear.

In previous work, we cloned and characterized a major portion of the gene for human von Willebrand factor. Presently we are applying these molecular tools to increase our understanding of the function of von Willebrand factor and the molecular basis for von Willebrand disease. von Willebrand factor is a large molecule with multiple discrete functions. We are conducting experiments to localize the various functions of von Willebrand factor to precise regions within the molecule. The specific segment involved in binding to factor VIII has been localized, and we are characterizing the regions required for von Willebrand factor interaction with the platelet surface and the blood vessel wall.

We are also focusing on the molecular basis of human von Willebrand disease. The defect in some patients with type III von Willebrand disease (the most severe form) is an inability to copy the von Willebrand factor gene into normal messenger RNA. Most of the patients with type IIB von Willebrand disease appeared to have one of four defects, all clustered within a small region of the von Willebrand factor gene thought to be 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 von Willebrand disease, we identified a separate set of defects clustered in a different region of the von Willebrand factor gene. By introducing these defects into tissue culture cells, we have shown that type IIA may be due to abnormalities in the process whereby von Willebrand factor is manufactured inside the cell. Finally, in studies of a patient whose von Willebrand factor is unable to bind factor VIII, we identified a single change in the gene that has helped to pinpoint the region of von Willebrand factor responsible for carrying factor VIII.

Through these studies, we hope to expand our understanding of von Willebrand factor, to advance our ability to diagnose and classify von Willebrand disease accurately, and to improve the medical treatment for this common human disorder.

Plasminogen Activator Inhibitor-1

The fibrinolytic system is the body's mechanism for breaking down blood clots. This system is delicately balanced with the system that forms blood clots. A precise balance between these two systems is vital. Overactivity of either system could result in uncontrolled bleeding or uncontrolled blood clotting. Plasminogen activator is the protein that turns on the fibrinolytic system. Its activity is controlled by the regulator protein —plasminogen activator inhibitor-1.

Synthetic plasminogen activators 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 has suddenly become blocked. There is also increasing evidence that patients with abnormally high blood levels of plasminogen activator inhibitor-1 (blocking the normal blood 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 plasminogen activator inhibitor-1 and of its interaction with plasminogen activators is of great significance in the study of the fibrinolytic system and its role in many important human diseases.

For these reasons we have cloned the gene for human plasminogen activator inhibitor-1 and have characterized its structure. We have successfully produced recombinant human plasminogen activator inhibitor-1 in bacterial cells and are using this system to explore its interactions with other components of the clotting system and to determine which portions are most important for its functions. This work should lead to an improved understanding of the role of the fibrinolytic system in human diseases and to the development of new treatment approaches for these important disorders.

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 bone marrow and would ordinarily be fatal. The patient is then "rescued" by the transplantation of normal bone marrow from a healthy donor.

We have used DNA probes to study the patient's blood cells after the transplant, to determine which cells have come from the donor bone marrow and which have come from any remaining patient bone marrow that survived the procedure. With recent advances we have been able to extend this approach to study the origin of even a single blood cell. In addition to increasing our understanding of the biology of bone marrow transplantation, this information may prove useful in directing treatment after the transplant and may help predict which patients are at risk for specific complications.

Molecular Genetics of Hemophilia A

Jane M. Gitschier, Ph.D.—Assistant Investigator

Dr. Gitschier is also Assistant 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, Inc., before joining the faculty at the University of California.

THE genetic basis of hemophilia has been understood since biblical times, when mothers of boys who had bled to death at circumcision were advised not to have future sons circumcised. The most common form of this disease, hemophilia A, affects 1 male in 5,000 throughout the world. It results from mutations in the gene coding for a blood coagulation protein called factor VIII. The factor VIII gene is located on the X chromosome, one of the two sex chromosomes. Consequently, hemophilia A is inherited as a "sex-linked" trait. Males, having a single X chromosome, are afflicted with hemophilia if their factor VIII gene is mutated, but females are protected from manifesting the disease by a second, normal X chromosome. Thus females may be silent carriers of the disease but transmit it genetically to their sons.

Our laboratory is interested in understanding what types of mutations lead to hemophilia A and how these mutations are generated. The disease 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. In addition, many cases of hemophilia occur sporadically in families without any history of the disease. This observation suggests that many mutations arise anew in affected individuals or their mothers. Knowledge of mutations allows accurate prenatal diagnosis and detection of female carriers by analysis of the mutation itself.

Because the factor VIII gene is unusually large, one of the major obstacles in this project lies in finding the mutations. We have experimented with several different mutation screens and have had considerable success with a technique called denaturing gradient gel electrophoresis. This technique is based on the principal that DNA fragments differing by a single base pair (e.g., normal vs. mutated) also differ in their thermal stability. In practice, the thermal stability of DNA fragments from normal and hemophilic individuals is assayed on gels containing a chemical denaturant, which has the same destabilizing effect as temperature.

We have focused on finding mutations in DNA samples from patients with mild and moderately severe hemophilia. Our experience suggested that these patients would be likely to have amino acid coding changes. We developed a strategy of screening for mutations in factor VIII coding regions by denaturing gradient gel electrophoresis. Mutations were found in 21 of the 35 patients screened to date. These mutations are scattered throughout the factor VIII gene and appear to be quite diverse.

Having identified the mutations responsible for hemophilia in some patients whose disease is sporadic, we have had the opportunity to investigate the origin of the mutations. In the past, mutations were assumed to occur as isolated events in either eggs or sperm. However, we have evidence that some mutations occur much earlier, at some point during parental development, rendering the parent "mosaic" for mutant and normal cells in different parts of the body. Consequently, many cells in the parent's germline may carry the mutation, but his or her blood cells, in which the presence of the mutation is tested, may not. Mosaicism has also been found associated with other X-linked and dominantly inherited disorders. These findings demonstrate that the unaffected parent of a child with sporadic disease may be at substantial risk of having a second child with the disease.

In a second line of research, our laboratory is exploring DNA sequences near the factor VIII gene. These are of great interest because many genetic diseases map near hemophilia A. For example, two of the most common X-linked diseases, red and green color blindness and glucose-6-phosphate dehydrogenase (G6PD) deficiency, are tightly linked to the hemophilia locus. By physical mapping methods, we found that the genes for factor VIII, visual pigment, and G6PD are located within a million base pairs of each other. In collaboration with Barbara Trask at Lawrence Livermore National Laboratory, we were able to establish the relative "orientation"

of the three genes, or the directions in which their messenger RNAs are transcribed.

A number of much rarer genetic disorders are also tightly linked to the factor VIII gene. These rare diseases are clinically diverse and of unknown biochemical basis. They include Emery-Dreifuss muscular dystrophy, nephrogenic diabetes insipidus, adrenoleukodystrophy, incontinentia pigmenti, and a manic-depressive disorder. By isolating genes near the factor VIII gene, we may be able to uncover the genetic basis for one of these disorders.

The physical map has formed the framework for identifying candidate disease genes in the factor VIII neighborhood. These are found by virtue of their association with "CpG islands," regions of human DNA that are rich in the bases cytosine and guanine. Since islands are often found near the origin of messenger RNA transcription, it is sometimes possible to identify and isolate a transcribed gene by isolating a CpG island. We have cloned five islands near the human factor VIII gene. Messenger RNAs have been found for three of the islands, and we have isolated cDNAs corresponding to two. These are being sequenced. Eventually the corresponding genes from patients will be examined for DNA sequence aberrations to determine whether the genes are responsible for disease.

A surprising result of the CpG island investigation was the discovery of an island inside a noncoding portion of the factor VIII gene. We found associated with the island two messenger RNAs, transcribed in opposite directions. The function of these "nested" genes is unknown, but their relationship to factor VIII is interesting to consider. Gene A is contained entirely within the noncoding region and is transcribed in the direction opposite to the factor VIII gene. Two other transcribed copies of gene A are located near the factor VIII gene. Their existence may explain why hemophiliacs deleted for the factor VIII gene have no disorders other than hemophilia. Gene B includes some of the factor VIII coding regions, suggesting that it may participate in the coagulation pathway. Currently we are trying to determine whether either of these genes produces a protein.

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.

SEVERAL years ago, while studying the replication of cells in culture, we discovered that compactin, a drug that can prevent the biosynthesis of a key metabolic intermediate called mevalonic acid, can also block the synthesis of DNA and prevent cells from attaching normally to a culture dish. Furthermore, we showed that compactin's effects on DNA synthesis and cell attachment depended on an induced deficiency of mevalonic acid. Both effects could be prevented by adding mevalonic acid to the cell culture medium, though known products of mevalonic acid metabolism, including cholesterol, were inactive. Thus it appeared that either mevalonic acid itself or an unknown product was involved.

In an attempt to identify the compound that was required for DNA synthesis and cell attachment, we treated cells with compactin in the presence of mevalonic acid with a radioactive tracer. Just enough exogenous mevalonic acid was used to prevent the drug's effects. Then we analyzed the radioactive products of mevalonic acid that the cells formed. This led to the discovery of a new class of modified proteins in animal cells.

As much as 40 percent of the cell-associated radioactivity was attached to a special group of cell proteins that could be distinguished by gel electrophoresis. The radioactivity seemed to be present in protein-attached products of mevalonic acid, called isoprenoid compounds. The discovery of this new type of modified proteins opened up a field of protein research that is now expanding rapidly.

In follow-up studies, we sought to identify the proteins and characterize the attached isoprenoid groups. We were able to show that one of the modified proteins is lamin B, a structural protein that is attached to the inner nuclear membrane. In collaboration with Michael Gelb (University of Washington), we demonstrated that lamin B is modified by a cysteine thioether-linked farnesyl group and that the farnesylated cysteine residue is located at the protein's carboxyl-terminal end. (A farnesyl group is a 15-carbon isoprenoid compound that is formed from three molecules of mevalonic acid.)

These results are of considerable interest, as other investigators have identified additional proteins, including the oncogene product p21 ras, that are similarly modified. Moreover, there is reason to believe that the modified carboxyl-terminal cysteine residues in lamin B and p21 ras promote binding of the respective proteins to cell membranes.

Some animal cell proteins are modified by farnesyl groups; other animal cell proteins are modified by another isoprenoid group, the 20-carbon geranylgeranyl moiety. In collaboration with Bernard Fung and Stephen Clarke (University of California, Los Angeles) and Masahito Kawata and Yoshimi Takai (Kobe University, Japan), we recently identified several geranylgeranylated proteins, including the γ-subunits of a brain heterotrimeric guanine nucleotide–binding protein and three low-molecular-weight guanine nucleotide–binding proteins—smg p21B, smg p25A, and G25K.

These results and those of other investigators have raised a number of important questions. Attempts to identify and characterize the enzymes that catalyze the various modification reactions are in progress in several laboratories. Work in our laboratory, done in collaboration with Fuyuhiko Tamanoi (the University of Chicago), has provided evidence that the farnesyl transferase activity of the yeast *Saccharomyces cerevisiae* depends on two different genes, *DPR1/Ram1* and *Ram2*, whereas the geranylgeranyl transferase activity depends on a third gene, *CDC43*. The precise role of these genes remains to be determined.

Another important question relates to the functional significance of the modification reactions. Many of the modified proteins appear to be attached to membranes, and the modifying isoprenoid groups appear to promote this attachment. But it is not clear how the modifying groups interact with the membranes at a molecular level. Experiments that address this question are under way.

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, where he studied with Elias Lazarides. 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, and the remaining cells 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 presently mutating *Dictyostelium* cells and isolating mutants with altered cell-type ratios. We will then use DNA transformation to identify the genes that were mutated. We are also using drugs that disrupt the cell cycle to examine the extent of the linkage between the cell cycle and cell-type choice.

Sensing Cell Density

During *Dictyostelium* development, a cell waits until it is near a large mass of other *Dictyostelium* cells before it turns on the stalk- or spore-specific genes. Being able to sense whether other cells are nearby represents a paradigm for possible mechanisms that would allow, for instance, liver cells 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 its 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.

We have found that *Dictyostelium* cells sense whether they are near other cells by secreting small quantities of a protein, which 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 cells diffuses away so quickly that it never reaches the threshold concentration. How-

ever, in or near an aggregate of around 1,000 or more cells, the DSF concentration will be above the threshold value. We have purified DSF and have found it modified by attachment of sugars. We are investigating whether the sugars are important for its function.

We have determined the amino acid sequence of four small fragments of DSF and have used this information to obtain part of the DSF gene. This has been used in turn to obtain the sequence of amino acids for part of the DSF protein. Computer comparison with data banks of other protein sequences indicates that DSF is not related to any known protein. We are now starting experiments to see if mammals have genes similar to the *Dictyostelium* DSF gene.

Molecular Genetics of Neuronal Recognition in Drosophila

Corey S. Goodman, Ph.D.—Investigator

Dr. Goodman is also Class of '33 Professor of Genetics and Neurobiology 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 with Nick Spitzer 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. We would like to uncover the molecules 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*.

Our ongoing cellular analysis of the developing central nervous system in the insect embryo has given rise to five major conclusions:

First, certain glial and mesectodermal cells appear to interact with one another, several undergoing specific cell migrations, and in so doing lay down a pattern for the major fiber pathways in the developing nervous system.

Second, these glial and mesectodermal cells provide instructive information for 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 appear to provide an attractive cue for the growth cones that extend toward the midline and pioneer the commissural axon pathways.

Third, once the 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 the nerve pathways are differentially labeled by recognition molecules that enable growth cones to navigate through complex choice points.

Fourth, our observations suggest that the expression of surface recognition molecules is dynamic and regional on the surface of individual neurons—that parts of the cell are differentially labeled in accordance with the processes around it for which it has a selective affinity. For example, a neuron might express different molecules on the surface of its commissural process as compared with its longitudinal process, reflecting its changing behavior as it navigates across the midline in a commissural pathway and then turns into a different longitudinal pathway.

Finally, having navigated along a series of pathways, growth cones are 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.

Our molecular genetic approach to these issues has been fourfold. First, we have been studying in *Drosophila* the expression and function of cell and substrate adhesion molecules that are likely to play a significant role in these events. For example, we cloned the three genes that encode the three subunits of *Drosophila* laminin, a substrate adhesion molecule previously shown to be a potent promoter of neurite outgrowth of developing vertebrate neurons. We have isolated a lethal mutation in the gene encoding the A subunit of laminin (*lama*) and are looking for mutations in the other two genes.

We have also cloned two *Drosophila* genes—*fat* and *dachsous*—that encode cadherin-like molecules (calcium-dependent cell adhesion molecules). One class of mutations in the *fat* gene cause a tumor-like overgrowth of epidermal tissues. Another class of *fat* mutations, as well as a class of *dachsous* mutations, alter the morphogenesis of epidermal tissues. Thus these two members of the cadherin gene superfamily are involved in the control of morphogenesis; at least one of the genes also functions as a tumor suppressor gene.

Second, beginning with an immunological approach, we identified and cloned the genes encoding four different surface glycoproteins. We call these proteins fasciclin I (fas I), fasciclin II

(fas II), fasciclin III (fas III), and neuroglian. These glycoproteins are dynamically expressed on overlapping subsets of axon fascicles and glia during embryonic development.

Three of these molecules, fas II, fas III, and neuroglian, are part of the immunoglobulin superfamily and are highly related to a family of adhesion molecules found on the surface of vertebrate neurons. The other protein, fas I, appears unrelated to any currently listed in the data bank. Our molecular genetic analysis has shown that, by a process of alternative RNA splicing, all four proteins are generated in a variety of different forms by different cells during development. To begin to test the function of these proteins, we used transformation techniques to induce their expression on the surface of the S2 *Drosophila* cell line in culture. With standard aggregation assays and biochemical methods, we have shown that these four proteins are homophilic adhesion molecules that can mediate selective cell sorting as well as cell aggregation; apparently fas I defines a new class of adhesion molecules.

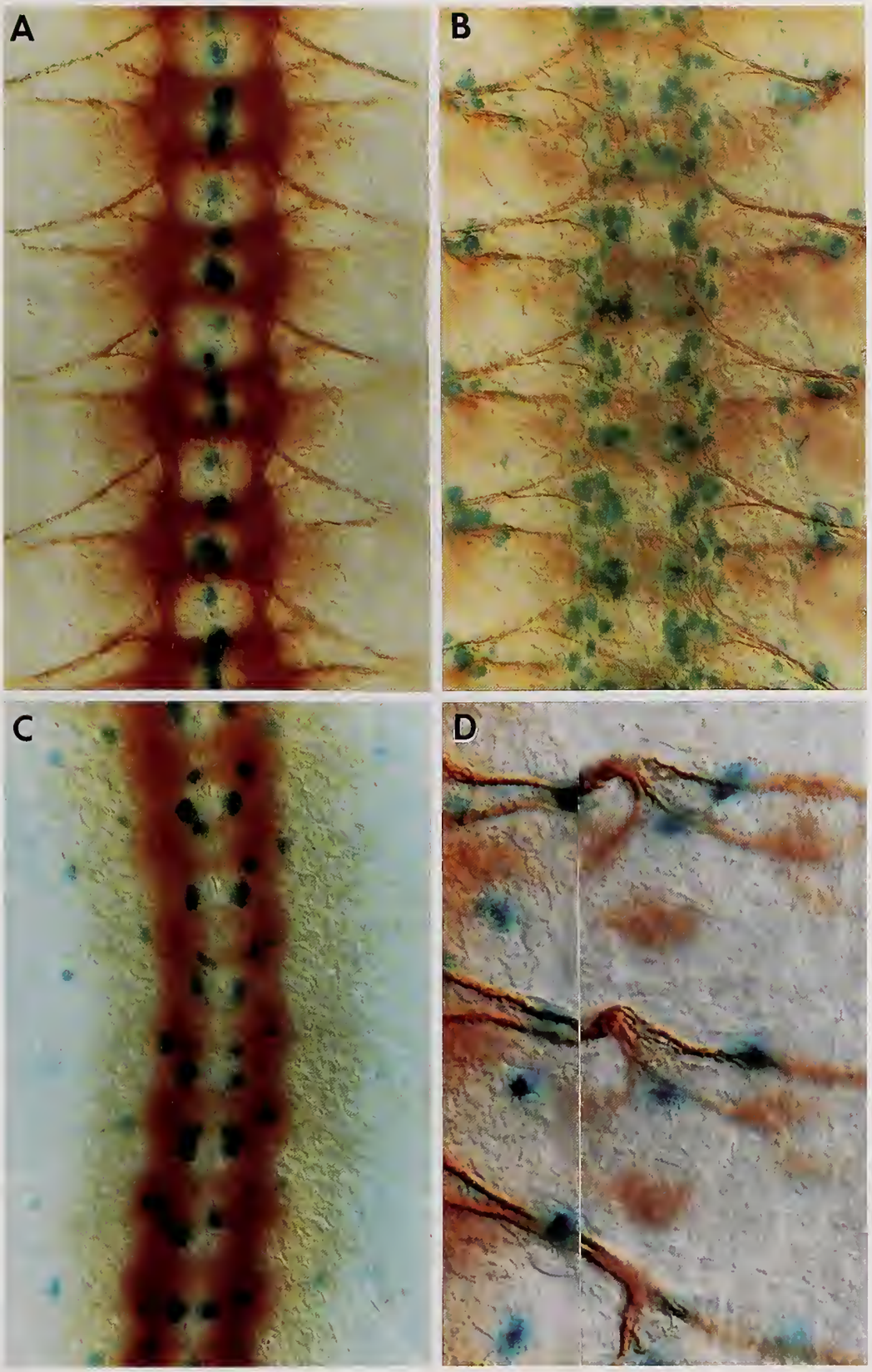
We are using genetic analysis and transformation methods to study the function of these molecules in the developing organism. We generated and identified insertional mutations and small deletions in the *fas I*, *fas II*, *fas III*, and *neuroglian* genes. Complete loss-of-function mutations in both *fas I* and *fas III* are viable under laboratory conditions; mutations in *fas II* and *neuroglian* lead to lethality. None of the four mutations lead to gross abnormalities in the developing nervous system. Rather, the mutations appear to lead to more subtle defects in specific subsets of neurons and specific axon pathways. For example, the fas II protein is normally expressed on a small subset of growth cones and axons that pioneer and selectively fasciculate in the MP1 axon fascicle. In the *fas II* mutant, these specific growth cones do not properly recognize one another and the MP1 pathway does not form; in contrast, other axon pathways appear to develop normally. Thus this member of the immunoglobulin superfamily functions in part as a neuronal recognition molecule for the MP1 axon fascicle.

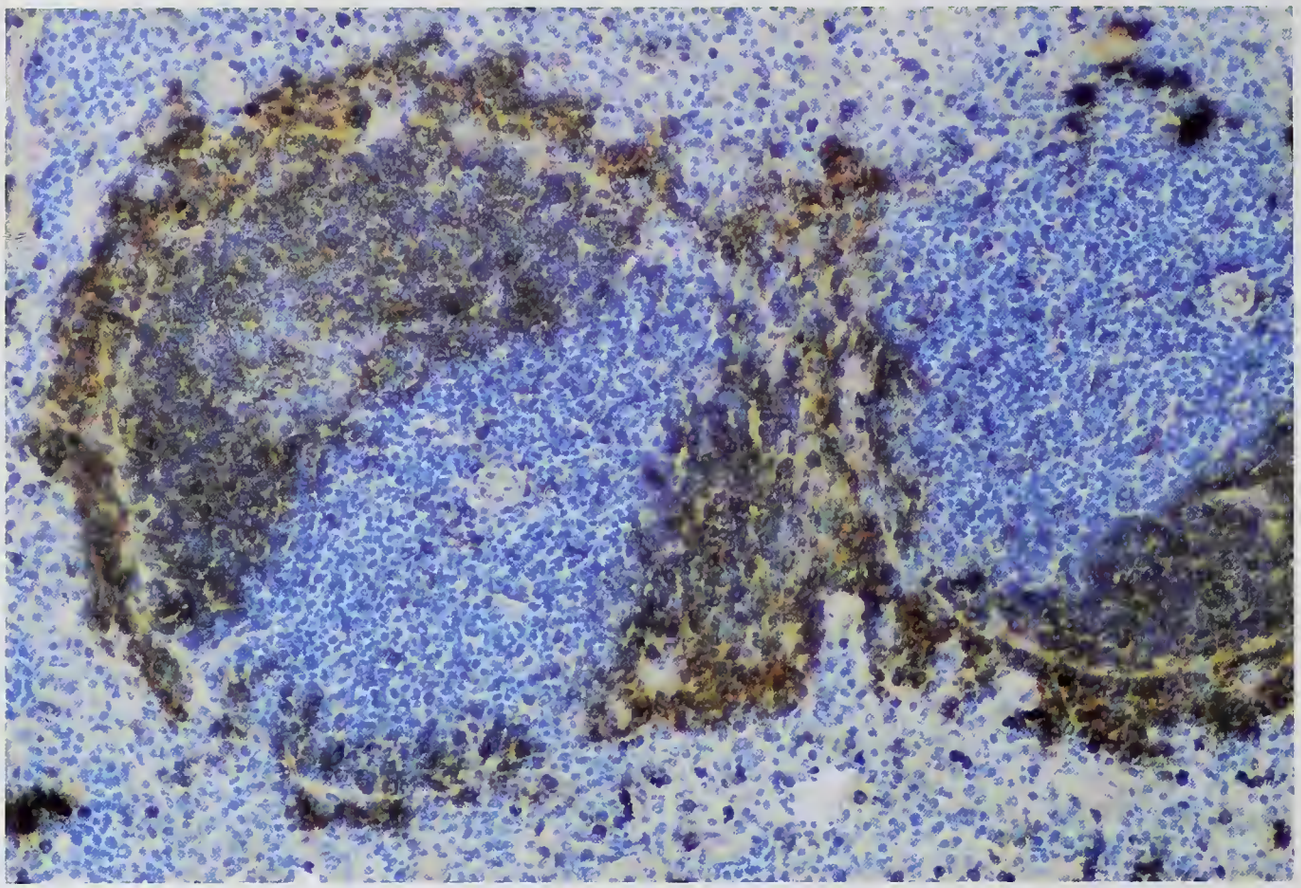
Third, we have been identifying and characterizing genes that function in the early interactions of the glia and mesectoderm in establishing the initial patterns of axon pathways or that function in the production or reception of the signals that attract the commissural growth cones toward the midline. For example, several new genes have been identified that appear to function in the guidance of growth cones toward the midline; mutations in these genes specifically disrupt the ability of commissural growth cones to extend toward and across the midline.

Fourth, we are using a variety of genetic and molecular genetic methods to screen for new genes that are involved in target recognition. As a model system for the study of target recognition in *Drosophila*, we focus on the ability of motoneuron growth cones to find and recognize their correct muscles during embryonic development. We have identified several genes that are expressed by different subsets of undifferentiated muscle fibers prior to innervation. Two of these genes encode membrane proteins that are members of the same gene family of cell adhesion and signaling molecules. We are presently characterizing mutations in these two genes.

Opposite: Visualization of axons (brown stain) and different subsets of glia (blue stain, revealing only nuclei) in or just outside the developing central nervous system in the embryo of the fruit fly Drosophila melanogaster. *The four panels show four different subsets of embryonic glia (as revealed by the "enhancer trap" method) in relation to the pattern of axon pathways that develop around them. Panel A shows the midline glia; panel B, the longitudinal glia and several other classes of nerve root and exit glia; and panel C, a small subset of these glia—all in relation to the pattern of axon pathways in four or five segments of the central nervous system. Panel D shows the four exit glia just outside the central nervous system in three segments.*

From Klämbt, C., and Goodman, C.S. 1991. GLIA *4:205–213.*


A
B
C
D



Immunohistochemical staining of B lymphocytes expressing transgene-encoded antigen receptors (brown), located within the follicular and marginal zone areas in the spleen of an immunoglobulin-gene transgenic mouse.

Research of David Y. Mason, Margaret Jones, and Christopher C. Goodnow.

Mechanisms of Immunological Self-Tolerance and Autoimmunity

Christopher C. Goodnow, 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 gene 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. In this particular case, where lysozyme is free in solution in the bloodstream, the B cells are indeed profoundly silenced, much as was predicted decades before.

Precise alterations in the design of the introduced cassettes are allowing us to build upon this model and explore the details of the normal process of self-tolerance. In one striking example, where lysozyme has been restructured so that it remains bound to cell membranes of transgenic animals rather than free in the blood, we find that lysozyme-binding B cells are not only functionally silenced but are in fact completely eliminated from the sites in lymphoid organs where immune responses are normally mounted. This finding points to important differences in the way the immune system tolerates different classes of self antigens in the body. We are planning to extend this approach to other important classes of self antigen by constructing additional strains of transgenic mice.

Another approach to understanding the process of self-tolerance in B lymphocytes makes use of the different forms of antibody, such as the immunoglobulins IgM, IgD, and IgG, which are differentially used as antigen receptors by B lymphocytes at distinct stages in their life span. By redesigning the antibody transgenes, we are currently comparing the behavior of B cells expressing these different receptor isotypes either

singly or in combination. These studies have begun to reveal a complex pattern of crosstalk between IgM and IgD receptors, which may help in dissecting the immediate events within the B cell that determine whether it is activated or rendered tolerant.

Once antigen binds to receptors on the B cell surface, the choice between being provoked into attack or becoming tolerant is ultimately determined by biochemical events inside each receptor-bearing cell. By studying the responses of tolerant and nontolerant B lymphocytes isolated from transgenic mice and maintained in tissue culture, we are beginning to identify some of the key changes involved in silencing the B cells that are reactive to self. These include both modulation of the antigen receptors themselves and blocking of the cells' ability to differentiate into antibody secretors. Transgenic mouse technology, allowing genetic dissection of signaling circuits within the B cells, together with standard techniques of biochemical analysis, should enable us first to gain an understanding of the normal controls exerted on B lymphocytes and eventually to illuminate the breakdown of these controls in autoimmunity.

Normal Human T Cell Growth, AIDS, and Adult T Cell Leukemia

Warner C. Greene, M.D., Ph.D.—Investigator

Dr. Greene is also Professor of Medicine at the Duke University Medical Center. He received his M.D. degree and Ph.D. degree in immunology from Washington University, St. Louis. After his internship and residency training in internal medicine at the Massachusetts General Hospital, Boston, Dr. Greene joined the Metabolism Branch of the National Cancer Institute at the NIH. Eight years later he was appointed to the staff at Duke.

THE normal T cell immune response defends the host from many infectious organisms (viruses, parasites, fungi, and bacteria) and eliminates cancer cells that may arise in the body. The catastrophic consequences of a defective T cell immune response are well illustrated by the acquired immune deficiency syndrome (AIDS), a fatal disease caused by infection of the helper subset of T lymphocytes with human immunodeficiency virus 1 (HIV-1).

Human T cell leukemia virus I (HTLV-I) also infects these helper T cells but, instead of producing T cell death, induces neoplastic transformation leading to adult T cell leukemia (ATL). HTLV-I has also been recently linked with a progressive demyelinating syndrome termed tropical spastic paraparesis, which resembles multiple sclerosis. This laboratory has focused its research efforts on further defining the biochemical processes that underlie both normal T cell growth and the abnormal patterns of proliferation that follow HIV-1 and HTLV-I retroviral infection.

The cascade of biological events constituting the normal T cell immune response is normally triggered by the interaction of foreign antigens with antigen-specific receptor molecules arrayed on the surface of the resting T cell. Antigen binding to these receptors in turn leads to T cell activation and the expression of various growth-related genes, including those for interleukin-2 (IL-2, T cell growth factor) and its high-affinity membrane receptor. The subsequent interaction of IL-2 with its receptor then promotes T cell proliferation, thereby increasing both the number and activity of the antigen-activated T cell clones.

Structure and Function of the Interleukin-2 Receptor

The functional IL-2 receptor corresponds to a membrane complex composed of two known ligand-binding proteins, IL-2Rα and IL-2Rβ. In the absence of its counterpart, the α- and β-subunits alone give rise only to low or intermediate receptor forms, respectively. The β-subunit plays a dominant role in growth signal transduction, while the inducible expression of the α-subunit gene is centrally involved in the tightly regulated display of high-affinity forms of the IL-2 receptor.

The IL-2Rβ gene is constitutively expressed in many resting T cells and natural killer cells. However, immune activation of these cells promotes augmented IL-2Rβ expression. Recent studies have revealed the noncovalent association of two different kinds of protein kinases with the IL-2Rβ subunit. Specifically, both a tyrosine kinase and a serine-threonine kinase interact with different segments of the intracytoplasmic domain of the IL-2Rβ receptor molecule. Studies are now under way to define the role played by kinases in growth signal transduction.

Transcriptional Activation of HIV-1

HIV-1 may establish a latent or persistent form of infection within host $CD4^+$ T cells and monocytes. Activation of these "latent" HIV-1 proviruses appears critically dependent upon select host transcription factors, including the NF-κB family of proteins. These inducible factors bind to and activate the duplicated κB enhancer element present in the HIV-1 long terminal repeat (LTR). Recent studies indicate that as many as four different κB-specific proteins are present in activated human T cells. Intriguingly, these four proteins are expressed in the nucleus with distinctly biphasic kinetics (two early and two late), a process that involves both differences in the mobilization of preformed pools of these proteins and changes in the levels of gene expression.

All four of these κB-specific binding proteins are structurally related to the v-*rel* oncogene from the avian reticuloendotheliosis virus, REV-T. Remarkably, wild-type but not nontransforming mutants of v-*rel* specifically bind to the κB enhancer and inhibit NF-κB function. These findings raise the possibility that v-*rel*-mediated cellular transformation may be inextricably linked to its suppressive interaction with the κB enhancer.

Trans-Regulation of HTLV-I Gene Expression

Our recent studies of HTLV-I have focused on the critical function of the Tax and Rex trans-

regulatory proteins encoded by this virus. The 40-kDa Tax polypeptide is a potent transcriptional activator that induces both the HTLV-I LTR and several cellular genes, including those for IL-2Rα and IL-2. The induction of these various transcription units does not involve the direct binding of Tax to target DNA sites, but rather is mediated through the indirect interplay of Tax with different sets of endogenous host transcription factors. Mutational studies of Tax have permitted the identification of discrete functional domains within this transactivator that are selectively responsible for the interactions with these different host factors. Furthermore, the amino-terminal 47 amino acids comprise two physically overlapping but functionally distinct domains that correspond to a novel zinc-binding motif and nuclear targeting signal. These domains both appear required for the biological function of Tax.

The HTLV-I Rex protein serves as an essential post-transcriptional regulator of HTLV-I structural gene expression, inducing *gag*, *pol*, and *env*. The structural and enzymatic products are uniquely translated from unspliced (*gag*, *pol*) or singly spliced (*env*) viral mRNAs, and are required for the assembly of fully infectious virions. Recent studies have shown that Rex action is mediated through a Rex response element, a complex RNA stem-loop structure located within the 3′ retroviral LTR. Mutational analysis of the *rex* gene has revealed the presence of at least two functional domains, one that mediates direct binding to its RNA response element and nuclear/nucleolar localization and a second that is required for Rex activation. Rex appears to function either by promoting the disassembly of spliceosomes engaging these long *gag*, *pol*, and *env* mRNAs or by activating the nuclear export of these incompletely spliced viral mRNAs.

Unexpectedly, the HTLV-I Rex protein is capable of functionally replacing the HIV-1 Rev protein, acting through the distinct HIV-1 Rev RNA response element. However, consistent with their lack of amino acid homology, Rex binds at a site in the HIV-1 RNA response element clearly distinct from that utilized by Rev. These findings highlight a pattern of convergent evolution in these two human retroviruses, whereby each employs a strikingly similar post-transcriptional strategy for regulating the expression of its structural gene products.

Additionally, the HTLV-I Rex response element plays an important role in the polyadenylation of all HTLV-I transcripts, independent of the presence of Rex. Specifically, folding of the RNA Rex response element serves to juxtapose upstream and downstream binding sites for essential nuclei polyadenylation factors, which must interact in a cooperative fashion to commit the HTLV-I site to effective 3′ processing. Alterations of the intrinsic secondary structure of the interposed Rex response element lead to the complete loss of 3′ end formation and correspondingly the abrogation of viral replication.

Developmental Control of Gene Expression

Rudolf Grosschedl, Ph.D.—Assistant Investigator

Dr. Grosschedl is also Assistant 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 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 Ig 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.

Regulation of Ig Gene Expression in a Transgenic Model

Our general experimental approach is to introduce intact rearranged Ig genes into cultured lymphoid and nonlymphoid cells and into the germline of mice. Comparison of transcription of the exogenous wild-type and mutant genes in different cell types and at various stages of development allows the identification of DNA sequences that are instrumental for tissue-specific and temporal gene regulation.

Transfer of a rearranged wild-type μ gene into the mouse germline indicated that the tissue-specific expression pattern is more complex than anticipated from previous observations. Instead of a simple "on-state" in lymphoid tissues and "off-state" in all nonlymphoid tissues, expression of both the endogenous μ locus and the μ transgene was also detected in skeletal muscle. Moreover, the μ transgene was reproducibly expressed at a low level in several other nonlymphoid tissues. Analysis of the expression pattern of mutant μ transgenes containing intragenic deletions or point mutations in nuclear factor-binding sites indicated a control by multiple positive and negative regulatory mechanisms.

Consistent with previous transfection data, μ transgene expression in lymphoid tissues was dependent upon the intragenic enhancer and the binding site for the Oct-transcription factors in the promoter. Neither of these regulatory sequences, however, was required for expression in skeletal muscle, which is governed by a muscle-specific control region located 3′ of the enhancer.

Low-level expression in other nonlymphoid tissues was independent of the Oct-binding site but requires function of the intragenic enhancer. Mutation of Es and E5 factor-binding sites in the enhancer increased this low-level expression to a high level similar to that found in lymphoid tissues. The off-state of the transgene in liver and fibroblasts, however, was not affected by this mutation, suggesting negative regulation by at least two mechanisms. Taken together, a plethora of regulatory information appears to be required for directing the complex tissue-specific expression pattern of the μ gene in the whole animal.

To determine the molecular basis for the sequential expression of Ig heavy- and light-chain genes during B cell differentiation, we generated chimeric Ig μ gene constructs, in which individual regulatory sequences have been replaced by the corresponding regulatory sequences from the κ light-chain gene. Gene transfer of these chimeric transgenes into the mouse germline and analysis of their temporal expression pattern indicated that replacement of the μ enhancer with the κ enhancer decreased the level of expression specifically in early-stage B cells and delayed, at least in part, the appearance of transgene transcripts.

Likewise, replacement of the μ promoter with the κ promoter decreased expression specifically in early-stage B cells. Neither the κ enhancer nor the κ promoter, however, was sufficient for conferring upon the μ gene the precise temporal expression pattern of the endogenous κ gene or an intact κ transgene, suggesting that the combination of the κ enhancer and κ promoter may be required for the proper temporal regulation of Ig genes.

Lineage- and Stage-Specific Expression of the *mb-1* Gene

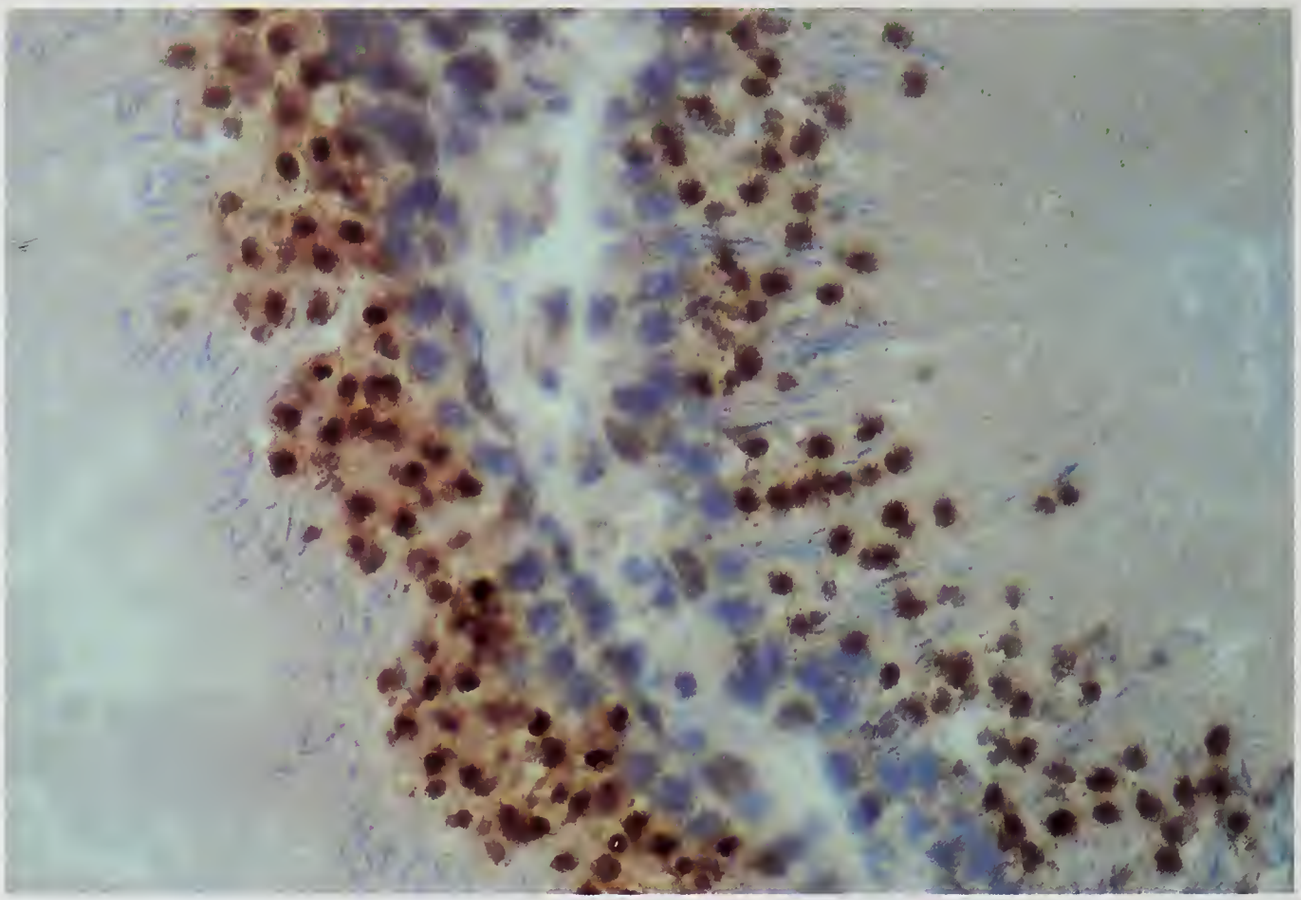
To gain insight into the molecular basis for the coordinate lineage-specificity and distinct stage-specificities of genes that are expressed only in lymphocytes, we extended our studies to the regulation of the *mb-1* gene. This gene is expressed specifically in early-stage B cells. Isolation and characterization of the *mb-1* promoter indicated that a relatively short DNA fragment contains the information for both lineage- and stage-specificity. The *mb-1* promoter was found to contain a binding site for a novel B cell–specific factor that is present only in nuclear extracts from early-stage B cells, thus paralleling the expression pattern of the *mb-1* gene.

Isolation and Characterization of Lymphoid-Specific cDNA Clones

To identify a set of new markers for individual stages of the lymphocyte lineage and to find gene products that are important for the lymphoid cell differentiation pathway, our laboratory isolated novel lymphocyte-specific cDNA clones by differential screening of a pre–B cell cDNA library with cDNA from an erythroid cell line. To identify candidate transcriptional regulators among the isolated lymphoid-specific cDNAs, we assayed proteins expressed from the cDNAs *in vitro* for binding to DNA cellulose.

One of the cDNA clones, which was expressed specifically in pre–B and T cell lines, was found to encode a 44-kDa protein that binds to DNA cellulose. Analysis of the nucleotide sequence of this cDNA revealed a homology to the nonhistone high-mobility group 1 (HMG1) gene and to genetically defined regulators of cell specifications that include fungal mating-type genes and the putative mammalian testis-determining factor gene *SRY*.

This lymphoid-specific cDNA clone represents a gene, termed *LEF-1* (*lymphoid enhancer-binding factor 1*), that is developmentally regulated and expressed in T and pre-B lymphocytes but not in later stage B cells or nonlymphoid tissues. *LEF-1* encodes a nuclear protein with a single HMG box that was shown to bind to a functionally important site in the T cell antigen receptor (TCR) α enhancer. Maximal TCRα enhancer activity was found to parallel the cell type-specific expression pattern of *LEF-1*. Moreover, forced expression of recombinant *LEF-1* in late-stage B cells increases TCRα enhancer function. Taken together, these data suggest that *LEF-1* is a regulatory participant in lymphocyte gene expression and differentiation.



Cyclic AMP–responsive enhancer-binding protein (CREB) expressed in germ cells of the testis. Shown is a photomicrograph of a cross-section of two seminiferous tubules at different developmental stages, VII (left) and V (right). The intense orange-brown structures are the nuclei of the round haploid spermatids stained with an antiserum to CREB. The relative acellular region is the interstitial space between the two tubules lined with the blue nuclei of the spermatogonia and spermatocytes, the cellular precursors of the spermatids. The round spermatids will differentiate into maturing spermatozoa, which are the linear blue structures adjacent to the spermatids (left) or juxtaposed between columns of spermatids (right). CREB is believed to activate the transcription of genes involved in spermatogenesis, under the regulation of cAMP-dependent protein kinase. Magnification, × 1,500.

Data of Gérard Waeber, Heather Hermann, and Joel F. Habener.

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. 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 the 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 our 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 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 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 a suppressor element prevents expression in the A and 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

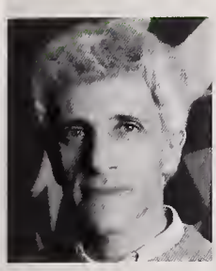
DNA-binding proteins and with adjacent target DNA sequences appear to determine cellular specificity of gene expression as well as metabolically regulated expression. Recently we have cloned several members of a family of cAMP-responsive enhancer-binding proteins (CREBs) and have discovered a domain 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 the profound effect of increasing the transactivation of gene transcription.

Discovery of an Insulinotropic Peptide

We have discovered two new glucagon-like peptides related in structure to pancreatic glucagon and co-encoded for glucagon in the prohormone (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 mRNA levels in insulinoma cell lines. We have determined that the glucagon-like peptide regulates insulin secretion in humans and may be involved in the pathogenesis of certain types of diabetes mellitus due to 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.

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 was recently elected to the National Academy of Sciences.

HOW do viruses enter and leave cells? How do receptors and their ligands cycle from cell surface to cell interior and back? How do regulatory proteins activate or inhibit transcription of particular genes? 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.

Viruses

The small double-stranded DNA viruses SV40 and polyoma have given us our 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 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. Flexibly extended arms, which form ordered structures only when the units assemble into a particle, appear to be an important feature of complex assemblies.

Comparison of a low-resolution structure of polyoma with the SV40 model shows that two surface loops are larger in the polyoma subunit. Mutational evidence suggests that these loops create 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 prepared cocrystals of polyoma with bound sialyl lactose in order to confirm assignment of its site and to study the molecular details.

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 formed 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.

We have also crystallized a rather different sort of cell-surface protein—the extracellular domain of the human transferrin receptor. This molecule undergoes a well-characterized cycle of uptake and return to the cell surface. The external domain, which makes up about three-fourths of the molecule, exhibits reversible conformational changes at low pH that we believe to be signals for intracellular sorting steps.

Transcriptional Regulatory Complexes

A common characteristic of eukaryotic transcriptional regulatory elements is the presence of sites in multiple copies that vary 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 com-

plexes containing the Cro protein of phage 434 or the DNA-binding domain of its repressor. We are beginning to use computational approaches to link observed structural differences among these various complexes with the corresponding free energies of binding.

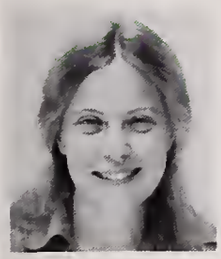
We have also been studying several eukaryotic regulatory proteins, initially by preparing crystals of their DNA-binding domains in complex with synthetic binding sites. TfIIIA, which controls 5S RNA transcription in *Xenopus*, represents the so-called zinc finger class. The finger is a small, 30-residue 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 (bp) DNA containing an appropriate part of the total binding site. We have crystallized this complex.

GAL4, a regulator of galactose metabolism in yeast, has a rather different zinc-containing structure. The amino-terminal 65 residues form a domain that binds with dyad symmetry to a 17-bp consensus DNA sequence. A first structure of such a complex is nearly complete. Each domain contains two zinc ions in close proximity, liganded by six cysteines.

GCN4, also a yeast regulatory protein, represents yet another class of DNA-binding structures. 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, and a structure determination is in progress.

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 components of a transcriptional initiation complex present even more challenging puzzles for the future.

Position-Dependent Gene Expression in Drosophila



Tulle I. Hazelrigg, Ph.D.—Assistant Investigator

Dr. Hazelrigg is also Assistant Professor of Biology at the University of Utah. She received her B.A. degree from Oberlin College and her Ph.D. degree in genetics from Indiana University, where she worked with Thomas Kaufman on the Drosophila Antennapedia *gene complex. She did postdoctoral work in the Carnegie Institution's Department of Embryology and in the Biochemistry Department at the University of California, Berkeley. In both places she worked with Gerald Rubin, analyzing DNA elements that regulate the expression of the* Drosophila white *gene.*

WE are studying two problems in the genetic regulation of pattern formation in *Drosophila*. One is the question of how determinants come to be localized within a cell (for instance, the egg) where they can initiate correct developmental fates. The second is the action of DNA regulatory sequences that respond to positional cues within a tissue and confine the expression of genes to particular cells.

During early development of the *Drosophila* embryo, maternal-effect genes express products involved in establishing the basic body plan. Among this class of genes is *exuperantia* (*exu*), which is needed for the correct determination of anterior embryonic structures. The product of the *exu* gene is needed during the development of the oocyte for anterior localization of the RNA of another gene, *bicoid* (*bcd*). This localization leads, during early embryogenesis, to a steep anteroposterior concentration gradient of *bcd* protein, which acts as an embryonic anterior determinant. We are interested in how the *exu* gene functions to bring about this subcellular localization of the *bcd* RNA.

We have analyzed the effects of *exu* mutations on spermatogenesis. During a long period of growth and development, primary spermatocytes, which have not yet undergone meiosis, appear normal in *exu* mutants; defects can be observed during meiosis and subsequent differentiation of spermatids. Fully differentiated, motile sperm are never produced. One *exu* allele is female-specific, showing only the maternal-effect phenotype, and the other allele is male-specific. Both alleles have been informative about the functioning of *exu* in gametogenesis in the two sexes.

We have cloned the *exu* gene and found that it encodes overlapping male and female transcripts, which differ in size in the two sexes. The two sex-specific transcripts differ also in splicing and polyadenylation patterns, but sequence analysis has shown that they both encode the same predicted 58-kDa polypeptide.

We have determined the sequence of the female-specific *exu* mutation. It changes a single amino acid in the protein, possibly identifying a region more critical for its female than its male function.

The male transcript is longer than the female in its 3′-untranslated tail. For part of the tail, the male-specific *exu* mutation is deleted. We are currently attempting to understand the function of the male-specific tail. We have also found that the *tra-2* (*transformer-2*) gene, which functions in *Drosophila* sex determination, is needed for efficient processing of the *exu* transcript in the male mode.

Together these results suggest that sex-specific processing of the *exu* transcript in the germline is a biologically important event. Why is the *exu* gene needed maternally for early development of the embryo and also during spermatogenesis? Since *exu* mutants disturb the anterior localization of *bcd* RNA in eggs, one model is that the gene functions in both cases to localize RNAs in the germ cells. Since there is no known effect of *bcd* mutations on spermatogenesis, different domains in the *exu* protein could act to recognize different RNAs.

Alternatively, the *exu* product could serve a more general role in germ cell cytoarchitecture, and the disruptive effect of *exu* mutations on *bcd* RNA localization could be a pleiotropic effect of disturbing this structure. We do not believe, however, that the latter scenario is correct. We have determined the distribution of the *exu* protein during oogenesis and early development with the use of antibodies raised against the protein. We find it to be present in the oocyte in a sharp concentration gradient at times in oogenesis when the *bcd* RNA is localized, with highest concentrations at the anterior ends of developing oocytes. But the protein appears to be degraded after it enters oocytes (it is produced by the nurse cells, which are attached to the oocyte and nourish it during its development) and is not present in the mature oocyte or early embryo. These results suggest that the *exu* protein's role in *bcd* RNA localization may be transient.

Perhaps the *exu* protein modifies another protein that the *bcd* RNA interacts with, or performs

an escort role to position *bcd* RNA in the right place in the oocyte. Another attractive hypothesis is that the *exu* protein may aid in folding the *bcd* RNA into a secondary structure that is recognized by other factors that localize it.

During development, positional information instructs genes to be expressed in different cells in given tissues. The *Drosophila white* gene is required for deposition of pigments in the eye of the fly. Normally the gene is expressed in the entire eye, leading to uniformly pigmented eyes in the adult. We are studying a case in which the *white* gene was inserted in a genomic location where its expression, rather than being uniform in the eye, is limited primarily to the ventral half. This patterned expression is due to the influence of DNA flanking the insertion site, since moving the gene to other locations in the genome does not have the same effect.

Mutations were induced that eliminate the patterned expression, returning the pigmentation to wild type. The molecular nature of these mutations has been studied by cloning and analyzing the flanking DNA. These studies show that sequences flanking both sides of the *white* gene are necessary for the regulation, but that a region upstream of the *white* promoter is particularly important for the effect. This regulatory region is sensitive to distance from the gene, since moving it closer causes more severe repression of *white* expression, leading to even less pigmented tissue in the eye. An intriguing aspect of the regulation is that it functions not only in cis to the *white* gene but also in trans on paired copies of *white*.

A number of *Drosophila* genes show pairing-dependent effects on their expression, suggesting that promoters or labile regulatory RNAs produced by these genes can act on a second homologous copy of the gene. We are analyzing the flanking DNA and unraveling its elements in order to understand better this proximity-dependent type of gene regulation.

Control of Gene Expression During the Cell Cycle and in the Developing Mammalian Cerebellum

Nathaniel Heintz, Ph.D.—Associate Investigator

Dr. Heintz is also Associate 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.

MOST interesting biological transitions, whether during the life of a single cell or the development of a complex tissue, are usually accompanied by underlying changes in the expression of genes. Knowledge of molecular events that result in activation of these genes can lead to a detailed understanding of such transitions. We are using a molecular and biochemical approach to examine specific transitions that occur in very different contexts: the mammalian cell division cycle and the developing mouse cerebellum.

Control of Gene Expression During the Cell Cycle

We have established over the past several years that coordinate induction of histone H4, H2b, and H1 gene expression during the S phase of the cell cycle is achieved through the agency of distinct transcription factors that interact with highly conserved subtype-specific consensus elements within each of the promoters. That the transcription of these genes is accomplished by different proteins suggests that their coordinate induction is due to a pleiotropic regulatory mechanism that directly participates in this activation step during the transition from G1 to S phase. Our efforts are focused on discovering the nature of this regulatory step and determining its specificity.

During the past year progress in elucidating this regulatory step has been achieved through in-depth analysis of the regulatory protein OTF1 (Oct1, NFIII) for the H2b cell cycle. Thus high-titer polyclonal and monoclonal antibodies specific for Oct1 have been raised and employed to examine the chemical nature of this protein as the division cycle progresses. Our results demonstrate that multiple forms of Oct1 exist in mammalian cells and that their distribution is dramatically regulated during the cell cycle.

Further analysis has established that differential phosphorylation of Oct1 during the cycle is at least partially responsible for the observed changes in this protein as cells proceed toward division. In particular, a peptide-mapping experiment using Oct1 pulse-labeled with ^{32}P orthophosphate *in vivo* has resulted in identification of at least six phosphopeptides that are specifically labeled in metaphase cells. Preliminary results indicate that purified Oct1 is a substrate for CDC2 kinase *in vitro*, but that these phosphorylations may be due to several protein kinases.

Our present efforts are focused on discovery of the enzymes that post-translationally modify Oct1 during the cell cycle, and demonstration that these modifications are functionally relevant to histone gene transcription. Similar analyses of a second histone-specific cell cycle regulatory factor, H1TF2, have begun to determine whether its properties are similarly modulated during the cycle. Demonstration that the timing and nature of the post-translational modifications on Oct1 and H1TF2 are similar *in vivo* would establish the existence of the proposed pleiotropic regulatory mechanism for regulation of transcription during S phase.

One fundamental question that has arisen from this work is whether the initiation of S phase–specific transcription and chromosomal DNA synthesis 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 chromosomal DNA synthesis? Might proteins that regulate initiation of 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 Medical School), we have recently identified a cellular protein complex (RIP60/RIP100) with several properties expected of replication-initiating proteins. Thus the purified complex binds specifically to two well-characterized origins of DNA replication at sites thought to be important for origin function, and shows ATP-dependent DNA helicase activity. Our present efforts are directed toward a definitive demonstration that RIP60/RIP100 can participate in DNA replication, and toward generation of appropriate tools to determine whether the ac-

tivity of these factors is regulated during the cell cycle.

Molecular Approaches Toward an Understanding 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 cerebellum 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 toward identification of these genes 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 RFLP (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. Initial results indicate that our cloning strategies have been successful, since many developmentally regulated cDNAs have been identified. For example, several candidate cDNA clones for the developmentally regulated neuron:glial ligand "astrotactin" have been obtained in a collaborative effort with Mary Beth Hatten (Columbia University).

Ribonuclease H and Genetic Replication

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 studies with Jerome Karle. He stayed on at NRL until he joined the faculty of Columbia University. His most recent honor is the Fritz Lipmann Award of the American Society for Biochemistry and Molecular Biology.

RNA-DNA hybrids have a vital role in biology as intermediates of genetic replication. DNA synthesis in both eukaryotes and bacteria is initiated by short RNA primers that are hybridized with the template DNA. This primer RNA must be removed for DNA replication to be completed. Various nucleases can perform this digestion, but notably specialized to the task are the enzymes known as ribonuclease H (RNase H, where H stands for hybrid), which hydrolyze only RNA chains and then only when they occur in heteroduplexes.

RNase H also participates in retroviral replication. It is a constituent activity of the multifunctional enzyme reverse transcriptase and, as such, is indispensable for retroviral infectivity. During reverse transcription, the polymerase moiety of the transcriptase molecule first produces an RNA-DNA hybrid on the template viral RNA. RNase H then removes the genomic RNA to free the complementary DNA, which serves as the template for plus-strand synthesis. Finally, the resulting DNA duplex is integrated into the host genome.

The RNase H domain constitutes the carboxyl-terminal third in reverse transcriptases of such retroviruses as human immunodeficiency virus (HIV) and Moloney murine leukemia virus (MMLV). These portions of sequence are clearly homologous with the RNases H of *Escherichia coli* and yeast. In light of our ultimate interest in reverse transcriptase structure, we have undertaken a crystallographic study of *E. coli* RNase H. This work, performed in collaboration with Robert Crouch, who discovered and extensively characterized this enzyme at the National Institutes of Health, proves to be fascinating in its own right.

MAD Structure of the Selenomethionyl Protein

We have used a novel method to solve the crystal structure of *E. coli* RNase H. This involved producing the recombinant protein with selenomethionine systematically replacing the four methionine residues in this 155-residue protein. The selenomethionyl protein is fully active, and it crystallizes isomorphously with the natural enzyme. We took these crystals, grown by Wei Yang, to the Photon Factory synchrotron in Japan. There, in collaboration with Yoshinori Satow, we made measurements of multiwavelength anomalous diffraction (MAD) for use in the structure determination. This MAD phasing analysis proceeded, in Wei Yang's hands, in a straightforward manner to an initial image at 2.2 Å resolution. Presently the structure is refined at 1.7 Å resolution with an R-value residual of 17 percent.

The RNase H from *E. coli* is an α/β protein with a tertiary folding that bears no significant similarity to other known structures. It is organized around a mixed central β-sheet of five strands with four α-helices on one side and one on the other. Our model is virtually identical with one reported independently by another group working with a different crystal form.

Implications for Substrate Binding and Catalysis

RNase H is an endonuclease that releases 5′-phosphate products upon hydrolysis of the RNA from RNA-DNA hybrid duplexes. The reaction requires divalent cations, preferentially Mg^{2+} in the case of *E. coli*. These and other features distinguish the catalytic mechanism of RNase H from that of the RNase A family. We can rather confidently identify the catalytic site in RNase H as being at the conjunction of seven evolutionarily conserved residues. At the center of this conserved group are a glutamate and two aspartate residues that site-directed mutations have shown to be absolutely required for activity. They form a carboxyl triad that appears to be a likely site for the binding of divalent cations.

Several structural features appear to mark the substrate binding zone. One is a sulfate-binding site, which we propose to be occupied by a nucleotide phosphate in the enzyme-substrate complex. This site lies in a groove separated by a ridge from the carboxyl triad, which is 14 Å away in another groove. The electrostatic potential surface of these grooves is complementary to that of

the nucleic acid, assuming that divalent cations are present at the carboxyl triad.

Based on these observations, we have built a model of the complex of RNase H with a hybrid duplex in the A conformation. We are attempting to cocrystallize the enzyme with oligonucleotide hybrids. We have also grown crystals of a complex with Mg^{2+} and are in the process of solving that structure. It appears that the catalytic mechanism is most likely to be similar to that of ribozymes and of the 3′→5′ exonuclease of DNA polymerase I, which Thomas Steitz (HHMI, Yale University) and his co-workers have studied in detail.

Relationships to Retroviral Reverse Transcriptase

The sequences of *E. coli* and yeast RNases H are clearly related to the RNase H portions of retroviral reverse transcriptases. Seven of the nine residues that are absolutely conserved among 13 retroviral RNases H occur in common with the microbial enzymes. Moreover, in a topologically restricted alignment, *E. coli* RNase H has 28 percent overall amino acid sequence identity with MMLV RNase H and 24 percent identity with HIV RNase H.

Our sequence alignments also suggest important structural differences among the various RNase H proteins. This has led us to an exercise in homology modeling in collaboration with Kenneth Smith and Barry Honig at Columbia. It appears that HIV RNase H has a 15-residue deletion relative to *E. coli* RNase H and that this can readily be accommodated. Our comparisons have also suggested mutations, which are now being studied in Steve Goff's laboratory at Columbia with the MMLV reverse transcriptase.

Finally we have returned to our structural studies on the MMLV reverse transcriptase. Millie Georgiadis is trying to produce usable crystals with new constructs prepared by Alice Telesnitsky in the Goff laboratory. We hope that what we have learned from *E. coli* RNase H, including MAD phasing based on selenomethionine, will help in our analysis of the retroviral enzyme.

Chromosome Organization and Gene Function 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 its expression. Thus most studies of gene expression are able to focus on the gene as an independent unit, without taking into account larger organizational features. There are exceptional cases, however, in which the relationship between a gene and its environment plays a role in expression of the gene. Our work has concentrated on two of these exceptions in the fruit fly.

Several years ago, we found a surprising association between two apparently unrelated genes. A gene encoding a secreted component of the *Drosophila* cuticle lay entirely within an intron of another gene, called *Gart*, which encodes enzymatic activities necessary for biosynthesis of purine bases. These oppositely oriented nested genes are regulated quite differently from each other during development. The cuticle gene belongs to a family of genes that is expressed at high levels in certain epidermal cells during two phases of the life cycle of the fly, late larval and prepupal. The *Gart* purine gene is apparently expressed at low levels in all cells throughout development, consistent with its housekeeping role. We wondered whether the expression of either gene was influenced by its neighbor—the cuticle gene by its surrounding neighbor, and the purine gene by its neighbor within.

Examination of the structure of the *Gart* locus in other insects revealed that the nested gene arrangement is nearly identical in a distantly related species of fruit fly (although the cuticle gene is absent from the intron of an even more distant non-*Drosophila* relative). The nested genes derived from one species are able to function correctly when introduced into the other species. By introducing segments of the locus from one species into the other, we are attempting to evaluate what components are necessary for expression of each of the nested genes, using the fully functional resident locus as a control.

Somewhat to our surprise, we found that regulation of either gene depends in part upon components closely associated with the other. These relationships between seemingly unrelated genes that occupy the same segments of DNA may have relevance to recently discovered examples of nested genes involved in human disease, including those for factor VIII and neurofibromatosis.

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-inacti-

vation 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. Our current efforts are aimed at identification of the cis- and trans-acting components of trans-inactivation.

Position effect in general and trans-inactivation in particular are phenomena that 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 in *Drosophila* will have general implications.

The classical phenomenon of dominant position-effect variegation reproduced at a site of insertion of the Drosophila brown *gene present on a transposon. The* brown *gene is necessary for transport of a precursor needed for red eye pigment production. The fly with a uniform red eye has a single copy of the transposed gene on one chromosome, while the other fly has both a normal copy and a position-affected copy. The mutant phenotype results from cis-inactivation of the position-affected copy and trans-inactivation of the paired copy.*

Research of Steven Henikoff.

The Evolution and Biological Roles 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, which was discovered nearly 100 years ago, was initially described as an activity found in serum that mediates the lysis of erythrocytes or the killing of foreign organisms also treated with specific immune antibodies. Later it was realized that complement also facilitates the interaction of antigen-antibody complexes with cells of the immune system. It is now known that complement can bind some foreign organisms or cells without the need for antibody.

The complement system consists of at least 20 serum proteins that are activated in a cascade fashion: initial activation of a small number of early components leads to the enzymatic generation of a large number of biologically active later components. As part of this process, protein fragments are released that attract inflammatory cells, and 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 proteolytically cleaved after attachment to targets. These cleavage reactions result in a number of different conformations; this allows C3 to interact with at least three unique cell surface receptors, the complement receptors. As part of this process, C3 fragments may also 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). CR2 interacts with the C3d,g fragment, which is located near the site at which C3 covalently binds to its target. The C3d,g fragment remains attached to targets after the rest of C3 is trimmed away during processing of the antigen-antibody-C3 complex.

CR2 also 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 of the human CR2 gene. Expression of the recombinant receptor in other cells is sufficient to mediate the binding of EBV and C3d,g. By using other recombinant techniques and creating mutations within the receptor, we have shown that specific amino acids in a small binding domain at the amino terminus of the receptor are important for ligand interactions. In addition, we have synthesized peptides that have the ability to block binding of EBV to the receptor. 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 viral binding but not normal binding of C3. This could be useful in some illnesses associated with EBV.

To increase our understanding of the biological role of complement receptors and regulatory proteins, we have cloned and expressed mouse homologues for CR2 and complement receptor 1, another C3 receptor. In addition, we have cloned and expressed related genes whose complement-binding and regulatory activities are not fully understood. Once we understand the activities of these proteins, we will be able to utilize many murine models of the normal immune response, as well as autoimmune diseases, to examine the roles that these proteins play *in vivo*.

Another aspect of CR2 expression is also under analysis in my laboratory. Expression of CR2 varies during 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 important in gene regulation. In addi-

tion, we are determining whether the receptor levels vary because the gene is turned on and off during development. 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 the Massachusetts General Hospital. 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 was recently elected to the National Academy of Sciences.

HOW do genes control animal development? Taking a primarily genetic approach to this question, the 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 many cell lineage genes is not constrained to a single tissue or organ. For example, one gene acts in both the nervous system and the musculature, while another acts in these two tissues as well as in the gonad and the hypodermis.

We have analyzed a number of these genes at the molecular level. These studies have revealed that genes that control aspects of 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. The study of this 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 may be of some 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. Although 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 decide which cells are to live and which are to die. We are currently analyzing molecularly genes involved in programmed cell deaths in *C. elegans*.

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 and are now characterizing genes that function in this signaling process.

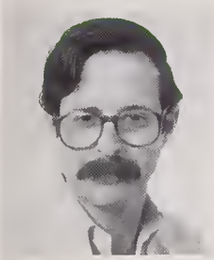
The second migration we have studied involves a pair of neuronal cells that move from the tail region to the midbody region of the animal. Thirteen genes have been identified that must function for these neuronal migrations to occur properly. Some of these genes probably act in the migration process *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.

Opposite: The microscopic nematode Caenorhabditis elegans *is used to identify and analyze genes that control development and behavior. Many of the genes that act in this simple animal are strikingly similar to genes found in humans. For example, the* C. elegans let-60 *gene is very similar to human* ras *genes. Mutations that cause* ras *genes to be abnormally active are oncogenic and constitute the most frequent class of mutation associated with human cancers.*

The figure shows a normal C. elegans *adult (top panel) and two different types of* let-60 *mutants. In the middle animal, the* let-60 *gene is inactivated, preventing certain cell interactions that normally induce formation of a component of the system used for egg laying. The animal is filled with retained eggs. By contrast, the lower animal is a mutant in which the* let-60 *gene is abnormally active, leading to the formation of abnormal growths (indicated by arrowheads). The study of the* C. elegans let-60 *gene, in providing insights to the function of* ras *genes, might help reveal ways in which* ras *gene activity and cancerous growth could be inhibited.*

Research of H. Robert Horvitz. Photographs by Greg J. Beitel.

Protein Folding In Vivo



Arthur 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.

WE have been using mitochondria, the intracellular organelles that carry out energy metabolism, as a system for investigating how proteins cross biological membranes.

Most of the proteins of mitochondria are first made outside the organelles, in the cytosol, and then imported through both an outer and inner mitochondrial membrane to reach the innermost "matrix" compartment. In order to traverse the membranes, the newly made proteins are first unfolded on the cytosolic side. After import, they refold on the inside of the organelle into their biologically active conformations.

It has always been assumed that this process of refolding is a spontaneous event, much like the refolding of many denatured proteins observable in a test tube. Yet we have identified a mutant cell in which mitochondrial proteins were imported into the matrix but failed to fold into biologically active forms. The mutation was found to affect a protein that normally resides in the matrix, called heat-shock protein 60 (hsp60).

This protein was originally identified by the observation that its level was increased about twofold in response to incubation of cells at high temperatures. It is abundant, however, 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 of the cell to efficiently refold 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. How does this complex function? 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 and a second protein component, the polypeptides are folded into their active forms and released from the complex.

We are now trying to dissect the mechanism of hsp60-directed folding. We believe that the pathway of folding 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 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 we can now reconstitute hsp60-mediated folding in a test tube, we can use biochemical and biophysical techniques to examine the folding mechanism more precisely.

How general is the utilization of folding machinery in living cells? Because mitochondria apparently arose from bacteria (one cell ingested another), it is not surprising that a structurally related component has been found in bacteria. Here a function like that of hsp60 is implied by an experiment carried out many years ago. When *Escherichia coli* cells partially defective in the hsp60-related protein were infected with virus, the newly made virus coat proteins could not assemble to make new virus particles. We surmise that the hsp60-related component is likely to be used not only to fold and assemble proteins of infecting viruses, but also to fold most if not all newly synthesized proteins indigenous to the bacterial cell. We are currently testing whether this is the case by producing a severe defect in the hsp60-related component.

Like bacteria, cells of higher organisms might utilize a "machine" to fold newly made proteins into their active forms. We are currently taking both genetic and biochemical approaches toward identifying such a component.

How does a folding machine like the hsp60 complex itself get assembled? It seemed possible that while all mitochondrial proteins tested to date utilize hsp60 for their folding and assembly,

the folding machine might self-assemble to produce a new complex. We recently tested this possibility and found that hsp60 subunits could not self-assemble inside mitochondria. Assembly required the presence of preexistent functional hsp60 complex.

How did the mitochondrial hsp60 complex get assembled in the first place? This question is answerable in terms of the origin of mitochondria. It seems likely that when the original mitochondrion was founded as an ingested bacterial cell, it already contained an hsp60-related protein. The question, then, really becomes, How was the hsp60-related component in the bacterial progenitor assembled for the first time? It seems possible that assembly might have occurred spontaneously, inaugurating a perpetual catalyzed assembly reaction. Alternatively, initial assembly might have occurred with the help of some other component—another protein, or perhaps a nucleic acid.

Thus we have shown that imported mitochondrial proteins do not fold and assemble on their own but, rather, require assistance from a molecular machinery. This may also prove to be the case for proteins of other cellular compartments. In general, in the living cell, newly made and newly translocated proteins may achieve their active conformations using machines that function like the hsp60 complex.

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, where he was Assistant Professor of Molecular and Cellular Neurobiology.

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 second neuron, thereby completing the process of synaptic transmission.

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 by extracellular factors. The molecular mechanisms that underlie this modulation 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 chosen the best-characterized neurotransmitter receptor and ion channel in neurobiology today, the nicotinic acetylcholine receptor, as a model system. In addition, we have been studying the major excitatory neurotransmitter receptor in the brain, the glutamate receptor, and the major inhibitory receptor in the brain, the $GABA_A$ receptor. 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; 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 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. Recently we have extended these studies to examine the effect of phosphorylation of the receptor on the β-, γ-, and δ-subunits by the protein tyrosine kinase. Phosphorylation of the receptor by the protein tyrosine kinase also regulates the rate of desensitization of the receptor. 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.

We have also 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, in collaboration with Gary Yellen (HHMI, the Johns Hopkins University School of Medicine). The desensitization and regulation of desensitization of these receptors by protein phosphorylation are now 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 appears to be secreted from neurons and to induce 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 of the glutamate receptor and the $GABA_A$ receptor have paralleled our studies of the nicotinic receptor. We are using protein purification, reconstitution, and site-specific mutagenesis techniques to determine the components required for neurotransmitter receptor function and to study the role of protein phosphorylation in the regulation of their function. These studies, combined with our studies of the nicotinic acetylcholine receptor, are likely to demonstrate 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 work 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 aspects of signal transduction processes responsible for vision. Vertebrate photoreceptor cells respond to a light flash via G protein–mediated activation of a cyclic GMP phosphodiesterase. A light flash hyperpolarizes a photoreceptor because hydrolysis of cGMP shuts down dependent plasma membrane cation channels. In darkness Ca^{2+} enters the cell through these channels, but a flash blocks its entry. The resulting decrease in cytosolic Ca^{2+} activates guanylate cyclase, which resynthesizes cGMP so that the cell recovers.

Invertebrate photoreceptors respond very differently. In these cells light activates phospholipase C, which produces inositol triphosphate and diacylglycerol as second messengers.

Much of our work focuses on the actions of G proteins, since many aspects of the visual response depend on this class of heterotrimeric signal transduction enzymes. G proteins bind GTP in response to receptor stimulation, and the GTP-charged α-subunit then dissociates from a complex of β- and γ-subunits to regulate the activity of an effector enzyme or channel. There are at least 15 different G protein α-subunits and 4 different types of β-subunits. This genetic diversity suggests that G proteins have a variety of functions and that G protein activities have been fine-tuned to generate specific cellular responses. The G proteins that mediate phototransduction in rod and cone photoreceptors are called transducins.

Vertebrate Phototransduction

Ca^{2+} mediates vertebrate rod cell recovery from a light flash. During the past year we identified a photoreceptor Ca^{2+}-binding protein, named recoverin, that activates guanylate cyclase only at free Ca^{2+} concentrations below 300 nM. Recoverin couples the light-induced loss of Ca^{2+} from the cell to guanylate cyclase activation and recovery from photoexcitation. In collaboration with Alexander Dizhoor, a visiting scientist from the USSR, we purified this protein, demonstrated its Ca^{2+}-binding properties, cloned and sequenced it, and demonstrated its ability to activate guanylate cyclase.

Once stimulated by light to bind GTP, transducin slowly loses its bound GTP and its ability to activate phosphodiesterase. Photoreceptor cells recover from a light flash within a couple of seconds, but the kinetics of GTP hydrolysis and phosphodiesterase deactivation are slower, about 20–30 seconds. To clarify the role that GTP hydrolysis plays in the photoresponse, we have produced transgenic mice that express a mutant transducin expected to hydrolyze GTP more slowly than its normal counterpart. Preliminary analyses suggest that the rod photoresponses from these mice are normal except in the presence of a high background light. Under those conditions, the cells appear to be abnormally sensitive.

Drosophila Vision

Biochemical and physiological evidence suggests that a G protein mediates phototransduction in the eyes of invertebrates by activating phospholipase C. We had previously characterized a *Drosophila* G protein β-subunit encoded by a gene referred to as *GBB*, and we expected to find this protein in *Drosophila* eyes. To our surprise, *in situ* hybridization studies revealed that GBB mRNA is absent from *Drosophila* photoreceptor cells.

We also found, however, that a monoclonal antibody raised against the GBB protein reacts with the entire *Drosophila* nervous system, including the eyes. In search of a protein in the eyes that reacts with this antibody, we used it to screen a *Drosophila* head cDNA expression library. The protein we identified is 50 percent identical to the *GBB* gene product, and the gene that encodes it, *GBE*, is expressed only in the eye. It is likely that *GBE* encodes the β-subunit of the *Drosophila* phototransduction G protein, but we need to do genetic and biochemical experiments to confirm this.

A *Drosophila* G protein α-subunit, DGOα, is absent from photoreceptors but present in the lamina, a layer within the *Drosophila* eye that

processes photoreceptor signals. The DGOα protein is a substrate for pertussis toxin, an ADP-ribosyl transferase that inactivates GOα subunits. We produced transgenic *Drosophila* that express a heat-shock-inducible pertussis toxin gene in most tissues of the fly, including the lamina. Pertussis toxin expression in these flies alters a transient component of the electroretinograms normally generated in the lamina. The photoreceptor component of the electroretinograms is normal, suggesting that pertussis toxin has uncoupled a normal photoreceptor response from its perception in the brain.

The long-term objective addressed in these projects is to identify the biochemical mechanisms by which cells determine their responses to stimuli in order to perform specific physiological functions.

Nectins and Integrins: The Molecular Basis of Cell-Substratum Adhesion

Richard O. Hynes, Ph.D.—Investigator

Dr. Hynes is also Professor of Biology 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 the 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, structure, 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 in order 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 the molecular analysis 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. We and others have analyzed 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. The interaction of these cells with these molecules is complex, 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 is one of the origins of the name "integrins."

We have recently made progress in analyses of the role of the intracellular portions of 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. Further progress along these lines should help explain the effects of cell adhesion on cell structure and behavior.

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. We have used recombinant DNA methods to produce specific and modified forms of fibronectins and integrins and investigate the ways in which they affect the behavior of individual cell types such as platelets.

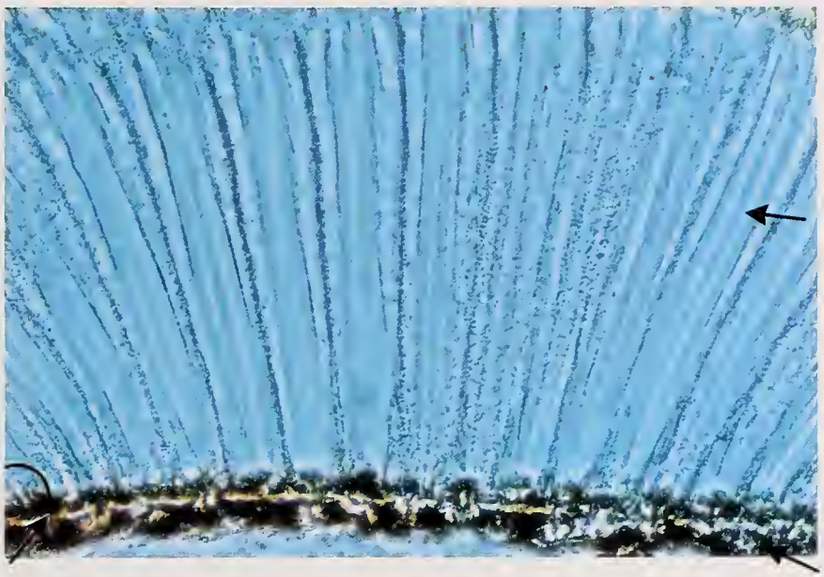
To extend our understanding of the roles of fibronectins and integrins in intact organisms, we use genetic analyses in two animal systems. We are analyzing the role of integrins during the development of *Drosophila melanogaster*, a fruit fly that is suitable for genetic analyses. Flies with mutations 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 insight into the functions of these proteins. We have also generated strains of mice that are mutant for fibronectins, and we plan to investigate the effects of these mutations on hemostasis, thrombosis, wound healing, and tumor development. We are also working to generate mice with mutations in integrins and in other molecules involved in cell adhesion.

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.

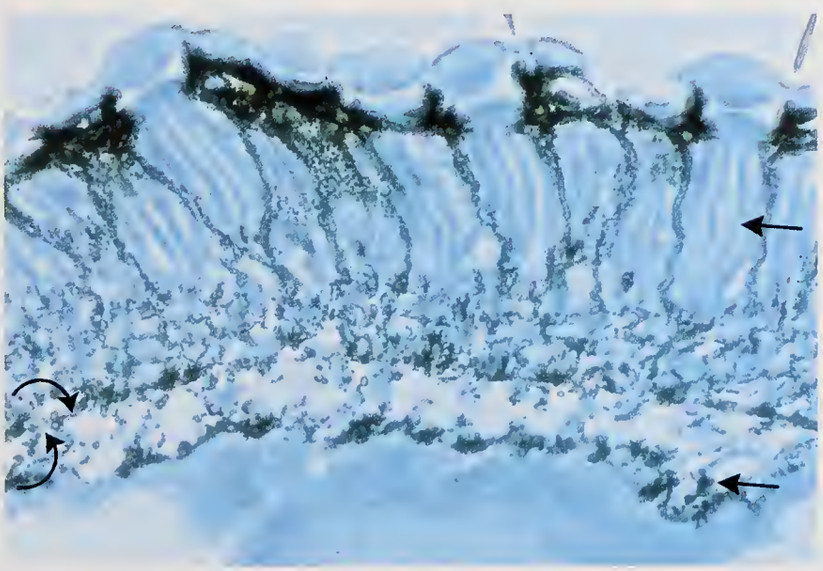
Opposite: Sections of the eyes of Drosophila melanogaster, *showing the regular array of photoreceptors attached to a basement membrane in normal (wild-type) retina and the disruption of this organization in eyes of flies with mutations in integrins.*

Research of Susan Zusman and Richard O. Hynes.

WILDTYPE RETINA



MUTANT RETINA



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.

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 invade 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 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. This microorganism is perhaps the most efficient bacterium at entering into human cells grown in culture, and this has greatly facilitated analysis.

To investigate intracellular growth, we have been analyzing *Legionella pneumophila*, the causative agent of Legionnaire's disease pneumonia. The intracellular growth process of this 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 103-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 the host cell's post-binding uptake requires only a 20-kDa region of invasin. Host cells can internalize a variety of bacterial species or even inert latex particles that are coated with this relatively small region of the protein. Evidently invasin's sole function is to present this binding region to host cell receptors in order to permit entry into the cell. 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. It is also well known that members of this receptor family interact in some functional fashion with an important host cell structure known as the cytoskeleton.

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.

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.

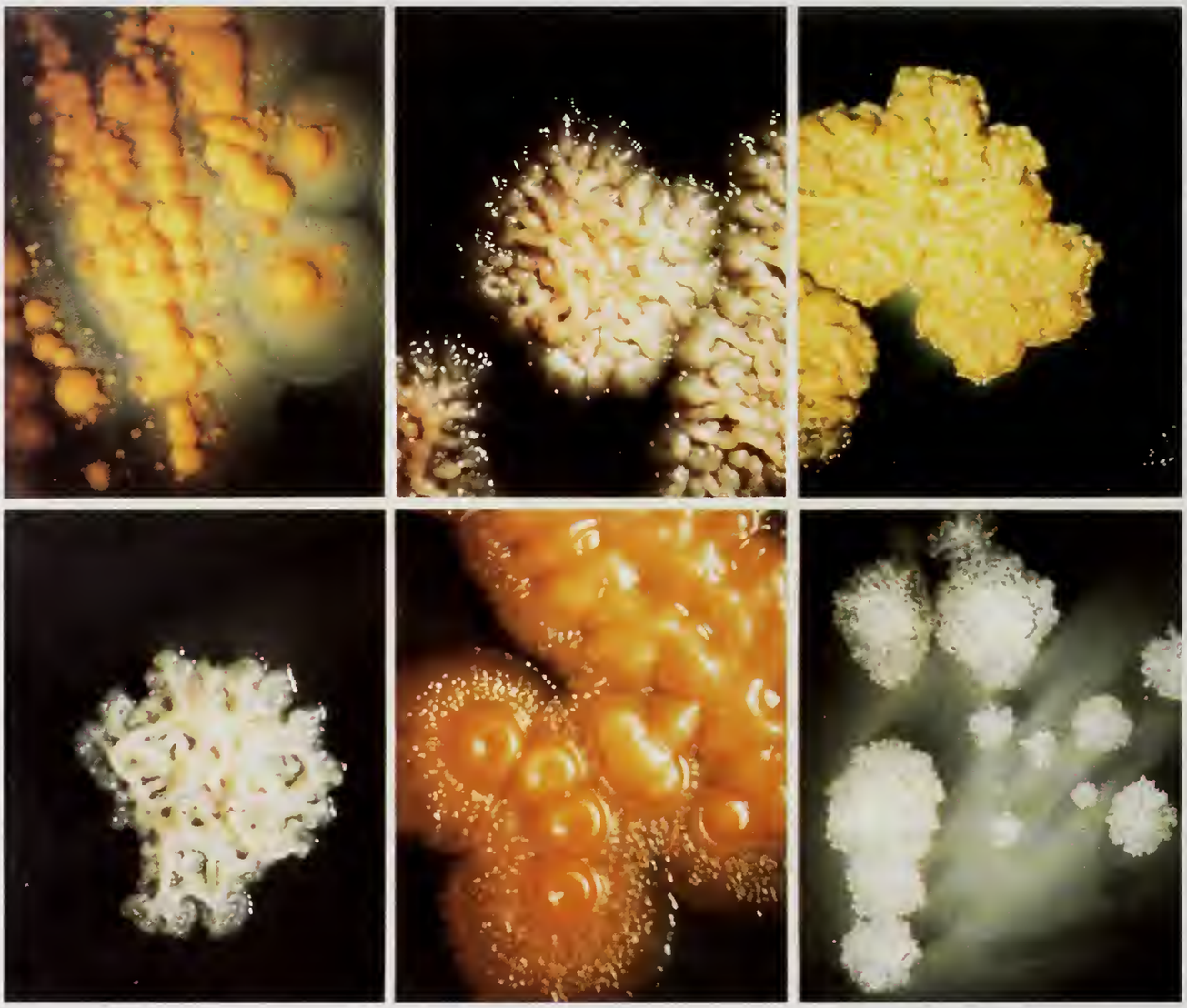
So, invasin appears to promote entry of the microorganism because it binds an important receptor that communicates with the cell cytoskeleton, and because it binds so tightly to this receptor.

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. Our approach has been to isolate mutations in this bacterium that prevent it from growing intracellularly. 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. We are currently trying to identify the factors missing in these mutants, using a combination of molecular and genetic techniques, in hopes of determining how the bacterium is able to perform each of these self-serving steps.



Colonial morphologies of a variety of fast-growing mycobacteria. Research of William R. Jacobs, Jr.

Genetic Approaches to the Control of Mycobacterial Disease

William R. Jacobs, Jr., Ph.D.—Assistant Investigator

Dr. Jacobs is also Assistant Professor in the Departments 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. His postdoctoral studies with Barry Bloom focused on developing genetic systems for the mycobacteria.*

TUBERCULOSIS, caused by *Mycobacterium tuberculosis*, continues to be the major cause of death throughout the world today. The World Health Organization estimates that each year there are approximately 10 million new cases of tuberculosis and over 3 million deaths. After 32 years of a steadily decreasing incidence in the United States, a surprising and alarming increase in the numbers of new cases has been reported in many of our cities in the last four years. This is thought to be a result of the AIDS (acquired immune deficiency syndrome) epidemic and is of considerable concern. Tuberculosis is not only a disease common to AIDS patients but is also one of the few diseases that can be readily spread from persons with AIDS to the general population.

Another microorganism of the same genus, *Mycobacterium avium*, although not a pathogen to healthy individuals, is a major opportunistic pathogen in AIDS. *Mycobacterium leprae* is the causative agent of leprosy, an affliction dating back to ancient times that affects over 13 million people in the world today.

In contrast to the pathogens, BCG (bacille Calmette-Guérin), the tuberculosis vaccine, has been used through parts of the world since 1922 to prevent tuberculosis. We believe that BCG represents an ideal candidate as a recombinant vaccine vector containing foreign antigen genes, because it is safe, has excellent adjuvant properties, and should elicit long-lasting immunity. Thus the goals of my laboratory are 1) to dissect pathogenic mycobacteria genetically in order to develop effective strategies to cure and prevent mycobacterial infections and 2) to engineer BCG as a recombinant vaccine vector.

Historically mycobacteria have played a prominent role in the development of microbiology. Robert Koch, in 1882, established the criteria (Koch's postulates) by which one ascertains whether an organism causes an infectious disease when he established that the tubercle bacillus causes tuberculosis. By analogy, "Koch's molecular postulates" is the method by which one ascertains that a characteristic of a bacterium, such as its virulence, is caused by a particular gene. The essential steps involve 1) identification of a mutant of the bacterium that lacks some characteristic, 2) cloning of an individual gene(s) from the parent bacterium, and 3) transfer of the cloned normal genes into the bacterial mutant to demonstrate restoration of the original characteristic. For example, to identify a gene required for virulence of *M. tuberculosis*, we would 1) identify an avirulent variant of the organism, 2) clone the genes from the virulent strain, and 3) introduce the putative virulence gene back into the avirulent mutant to demonstrate that this gene confers virulence.

For the mycobacteria, these sorts of experiments entail considerable difficulties. The organisms have been objects of study since Koch's pioneering work, but their genetic analyses had not been achieved, primarily because of their slow growth. The tubercle bacillus, which multiplies only once every 24 hours, requires three weeks to form a colony from a single cell. In contrast, *Escherichia coli* yield visible colonies in eight 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. Furthermore, although recombinant DNA technology has enabled us to clone any mycobacterial gene, the technologies to transfer recombinant DNA back into mycobacteria did not exist five years ago.

We took advantage of bacterial viruses that infect mycobacteria, called mycobacteriophages, as the building blocks for developing vectors capable of transferring recombinant DNA efficiently into mycobacterial cells. A novel hybrid vector, called a shuttle phasmid, was constructed. It replicates in mycobacteria as a phage and in *E. coli* as a plasmid, which permitted us to introduce cloned genes into a wide variety of mycobacterial strains for the first time.

Using these vectors, we identified selectable marker genes that have allowed us to develop phage- and plasmid-derived vectors, as well as systems to insert genes into the mycobacterial chromosomes. We have isolated mutants of a fast-

growing mycobacterium, *M. smegmatis*. These exhibit high plasmid transformation frequencies and thus provide a most useful surrogate host for analyzing the genes of slow-growing pathogenic mycobacteria.

Genetic Analysis of Mycobacterial Virulence Determinants

A major focus of this laboratory is to elucidate the causal mechanisms of mycobacterial diseases by identifying the genes responsible for specific virulence characteristics. The systems we have developed allow us to introduce libraries of genes from virulent mycobacteria into avirulent strains with great efficiency. A complementary strategy is the generation of specific mutants that no longer are virulent in our animal models. The combination of these strategies has allowed us to identify a number of genes necessary for the biosynthesis of amino acids, purines, and complex polysaccharides found on the surface of the pathogenic mycobacteria. By identifying the genes and their products responsible for virulence, we hope to be able to design approaches to the control of mycobacterial disease. In addition, the genetically engineered avirulent mutants should provide promising vaccine candidates.

Genetic Analysis of Drug Targets of the Pathogenic Mycobacteria

Although there are various effective antibiotics for tuberculosis and leprosy, antibiotic-resistant bacteria are continually emerging as a major public health problem here and abroad. Moreover, there is no effective chemotherapy for *M. avium* infections. We are employing genetic approaches to identify the mechanisms of action of present anti-mycobacterial antibiotics and the mechanisms of resistances to these antibiotics. The basic information so acquired should lead to the design of more effective drugs, strategies to overcome resistances, and diagnostic tests to detect mycobacterial strains containing resistance-conferring genes. We hope that improved strategies for treating leprosy, tuberculosis, and other mycobacterial infections can be developed.

Development of Recombinant BCG Vaccines

By cloning and introducing foreign antigen genes into BCG, it may be possible to develop novel vaccines that would protect not only against tuberculosis but also against the pathogen from which the foreign gene was obtained. In collaboration with Barry Bloom (HHMI, Albert Einstein College of Medicine) and colleagues at MedImmune and the University of Pittsburgh, we are developing systems to express foreign antigens from a variety of parasitic, viral, and bacterial pathogens. We have devoted considerable effort to elucidating the mycobacterial genetic signals necessary for expression and stable replication of foreign genes. Preliminary results demonstrate that high levels of foreign proteins can be produced in BCG and that these recombinant vaccines can elicit both humoral and cellular immune responses in infected animals. It is our hope that recombinant BCG will not only yield effective vaccines but also provide a novel and useful tool for studying immune responses in mammalian hosts.

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. Her Ph.D. degree was in physics and biophysics, mainly because she studied high-energy theoretical physics before becoming a biology student. After graduate school 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.

In order 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. Subsequent studies by Larry Salkoff, Mark Tanouye, Alberto Ferrus, C. F. Wu, and Leslie Timpe provided strong evidence suggesting that the *Shaker* locus codes for a component of a rapidly inactivating potassium channel, the A channel. 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.

These *Shaker* proteins appear to be integral membrane proteins, as indicated by their sequence and by subsequent immunoblot (Western) studies. They show homology to the sequence of vertebrate sodium channels, although they are much smaller in size. They correspond roughly to one of the four internally homologous domains of the sodium channel. Finally, Dr. Timpe demonstrated that RNA encoding for four of the *Shaker* proteins of known deduced sequence, when injected into frog oocytes, causes the functional expression of potassium channels that in several ways resemble the A channel in the fruit fly. Taken together, these studies showed that the *Shaker* locus is a potassium channel gene.

Starting with the *Shaker* gene in the fruit fly, Dr. 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. By now more than 10 different mammalian potassium channel genes have been characterized by a number of laboratories. One of these (the rat *Shal1* gene), cloned and characterized in our laboratory by Tim Baldwin, Meei-Ling Tsaur, and George Lopez, 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 between these mammalian 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.

An intriguing finding is that the *Shaker* potassium channel gene in the fruit fly gives rise to a number of protein products by alternative splicing of the primary RNA transcript. These different protein products probably form different subtypes of this potassium channel, because they show differential tissue distribution and, when expressed in frog oocytes, they give rise to channels of different kinetic properties. Alternative splicing of the potassium channel gene, together with the potential mix and match of subunits from one or more genes, offers a possible mechanism for generating potassium channel diversity.

Having cloned a potassium channel gene, 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.

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 not only will contribute to our understanding of channel function but also will be relevant clinically, for example, in the development of more specific drugs for 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.

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 the last few years, we have been interested in the following questions in neural development: How do neurons arise from undifferentiated ectodermal cells? What gives them their individual identity as to shape and function?

Our long-term goal is to understand these processes at the molecular level. Our approach, essentially genetic, is first to isolate mutations that affect neurogenesis, neuronal type, or axonal pathway formation, and then to identify these mutations, leading to the isolation of important genes.

During the last five years, our laboratory has been engaged in an extensive search and analysis of mutants affecting neural development in *Drosophila*. To identify and analyze such mutants, we are 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). This has resulted in the identification of a number of genes that specify cell fate in the embryonic fly. Analysis of those genes led us to propose a model for "progressive determination of the PNS."

Early during embryogenesis, cells in different locations within the ectodermal layer acquire unequal developmental potential, as a result of genes that specify positions in the embryo, the "prepattern genes." These include genes that specify dorsal-ventral and anterior-posterior axes. Some cells apparently acquire the potential to become neuronal precursors from the action of "proneural genes," which include, for example, genes of the *achaete-scute* complex (*AS-C*) and *daughterless* (*da*). Both *AS-C* and *da* encode proteins with the helix-loop-helix (HLH) structural motif. It is likely that *AS-C* and *da* products form homo- or heterodimers that bind to DNA and regulate the transcription of target genes, initiating neuronal precursor development.

As a neuronal precursor forms, it inhibits neighboring cells from assuming this role. Such "lateral inhibition" involves the action of "neurogenic genes." Removing the function of any of the six known zygotic neurogenic genes—*Notch* (*N*), *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 *N* and *Dl*, both encoding for membrane proteins with epidermal growth factor (EGF)-like repeats. *E(spl)-C*, *mam*, and perhaps *neu* are involved. 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 *E. coli* glycerol facilitator, which allows passive transport of small molecules such as glycine. The commitment of neuronal precursors may involve the actions of certain "master regulatory genes," which lock a cell into a particular fate.

The identity of a neuronal precursor is further specified by "neuronal type selector genes." These control the type of sensory neuron that a precursor will give rise to. For example, the *cut* locus is required for external sensory organs to acquire their correct identity. In the absence of *cut* function, those organs are transformed into chordotonal organs. Normally *cut* is expressed in external sensory organ precursors but not in chordotonal organ precursors. We think that expression of *cut* gene activity determines which of the two types of organs the precursor will become. The *cut* product contains a homeodomain and probably acts as a transcription factor regulating the expression of downstream differentiation genes.

Sequence information on a number of genes involved in neural development indicates that the majority of these genes contain a previously identified functional motif—e.g., the EGF repeat in *Notch*, the tyrosine kinase domain in *seven-*

less, the homeodomain in *cut*, and the helix-loop-helix motif in *da*. In each case, the motif immediately suggests 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 messengers, and regulators of gene expression. Perhaps only a modest number of new tools had to be invented for the formation of the nervous system, and many of the cellular mechanisms in neural development may turn out to be ones already in use before the nervous system evolved. Understanding neural development may require an understanding of the usage and manipulations of these basic functional motifs.

It is apparent that many of the functional motifs have been highly conserved during evolution. Several hundred million years separate the fly and vertebrates, yet the aforementioned motifs—EGF repeat, tyrosine kinase domain, etc.—are clearly conserved. Thus studying development in organisms with well-developed genetics, such as fruit fly and nematode, may provide not only reasonably satisfactory answers but also useful tools for the study of corresponding problems in vertebrates.

The Discrimination of Infectious Non-Self from Noninfectious Self

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 the NIH and the Biomedical Center in Uppsala, Sweden. In 1990 the students and faculty of Yale University Medical School selected Dr. Janeway as the Outstanding Teacher of Basic Biomedical Sciences.

THE key function of the immune system is to recognize infectious agents as foreign and eliminate them from the body. The defense mechanisms for the elimination phase can also damage the body's own tissues. This is normally avoided through a process of self/non-self discrimination mediated by highly specific receptors for antigen. Our studies have focused on two aspects of this process. First, we have explored the nature of receptors on T cells and of ligands on self cells in both normal immune responses and experimental autoimmune diseases. Second, we have probed the mechanisms by which the immune system identifies a foreign material as infectious. A main objective is to identify evolutionarily stable microbial constituents that act on nonclonally distributed receptors and lead to the induction of signals alerting the immune system to the presence of foreign matter.

We have examined activation of normal T cells that requires both binding of ligand to their clonally distributed receptor and delivery of "co-stimulatory" signals through a still uncharacterized receptor-ligand mechanism. These studies have also revealed a diverse collection of microbial constituents that can induce the co-stimulator required for clonal expansion of T cells. By means of two signals—a specific ligand recognized by the clonally distributed receptor, and a nonspecific signal derived from a host cell upon activation by a stable microbial constituent—the immune system can discriminate infectious non-self from noninfectious self. Without the delivery of the co-stimulatory factor, no clonal expansion or differentiation to effector function occurs. Thus a major component of self/non-self discrimination is mediated by the system of nonclonally distributed receptors for stable microbial constituents.

Immune responses are triggered by specialized antigen-presenting cells whose distinguishing feature is the ability to express co-stimulatory factors. Immune responses, however, must be able to detect infection in normal tissue cells that do not express co-stimulatory signals. For this to occur, the delivery of effector responses must not be dependent upon co-stimulation—and is known not to be in the case of cytolytic T cells and helper T cells. A risk, however, is inherent in this aspect of immune effector function: the occurrence of T cell responses to foreign microbial agents that cross-react with host cells. Such responses might lead to autodestruction of host tissue.

Avoidance of such destruction, however, might be effected through a third mechanism that we have recently revealed. Activated cells that encounter ligand in the absence of co-stimulatory factors, as on the surface of host tissue cells, are induced to secrete effector lymphokines and then abruptly die. The mechanism of death involves interferon-γ produced by the effector T cell, and perhaps inhibited by the cytokine interleukin-4. This process of activation-induced cell death has been observed both in cultured cells and, recently in other laboratories, *in vivo*.

If self/non-self discrimination is to be regulated not only at the level of receptor-ligand interaction but also by delivery of co-stimulatory signals, as required by all peripheral mechanisms of tolerance, then there must also be a requirement for both the specific ligand and the co-stimulator to be delivered by a single cell. This would be necessary to avoid the activation of T cells that recognize ligand on the surface of a tissue cell but receive co-stimulation from resident tissue antigen-presenting cells. We have examined this requirement for normal CD4 T cells and find that it does indeed apply. That is, delivery of the ligand for the T cell receptor by one cell and separate delivery of the co-stimulatory stimulus does not lead to T cell activation, clonal expansion, or the acquisition of effector function. This requirement allows antigen-presenting cells to be distributed throughout our tissues, where they can acquire local viruses and lead to the induction of anti-viral immune responses without jeopardizing the critical parameter of self-tolerance.

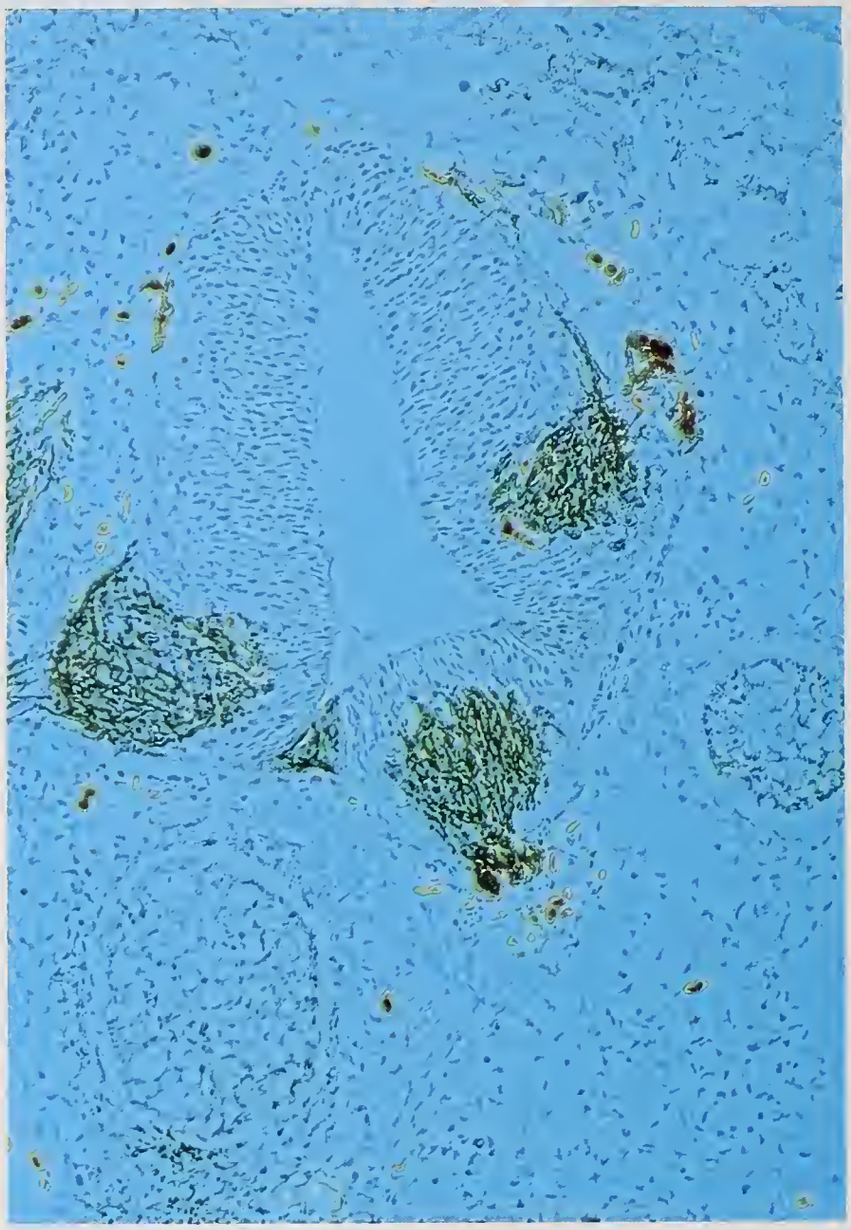
Given all of these restraints on self reactivity, one might ask how autoimmunity can ever occur? One mechanism that we have uncovered operates through the B lymphocyte. Activated T cells recognizing antigen on the surface of a B lymphocyte

can induce the expression of co-stimulatory activity in the latter cell. It has been known for some time that T cells reactive to a foreign epitope linked to a self protein can activate B cells that will produce antibody directed at the protein. Our studies have now shown that anti-self protein B cells become activated in this way so that they can now present the self protein in an immunogenic form to self T cells, leading to a sustained autoimmune T cell response. We have generated autoreactive T lymphocytes by this means and are testing whether they can generate a sustained autoimmune response *in vivo*.

Now that the phenomenology of discrimination of infectious non-self from noninfectious self has been established, and experimental systems put in place to characterize this response, our studies will begin to focus on the receptor-ligand pairs involved in the discrimination of infectious organisms from the noninfectious host. A system already established to detect the increase in co-stimulatory activity has revealed that inducers of co-stimulation act on virtually all antigen-presenting cells. Assuming that this process operates through a receptor-ligand pair, the result demonstrates that such receptors are nonclonally distributed, as we predicted in 1989.

We propose that the receptors currently utilized by the immune system to discriminate infectious from noninfectious are those that once allowed primitive organisms to regulate their immune effector responses. Such receptors may still function in the early nonclonal phases of host defense, as well as in triggering the induction of specific clonal defense as discussed above. If this proves to be so, the innate and specific immune responses will be shown to be linked not only at the effector phase but also at the recognitive phase. Such a synthesis is the long-term goal of our research in this area.

On a more practical level, identification of receptors that regulate the induction of immunity may lead to up- and down-regulation of immune responses to achieve the desired level of effector function in such situations as allergy, autoimmunity, graft rejection, and cancer.



Changes in neural cell pattern induced by the notochord and floor plate. Micrograph shows the change in pattern of floor plate and motor neuron cell differentiation in response to a grafted notochord in chick embryos. In this section of an embryonic chick spinal cord, the left side shows a normal ventral motor neuron pool, labeled with a monoclonal antibody. A grafted notochord adjacent to the right side of the spinal cord induces an additional floor plate and ectopic motor neurons in dorsal regions.

Research of Thomas M. Jessell.

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.

A major aim of our research is to define the mechanisms that control the patterning of neural tissues in vertebrate embryos. Our studies continue to focus on the events contributing to the differentiation of specific cell types in the developing neural tube. Over the past year we have obtained evidence that the floor plate, a specialized group of neuroepithelial cells, is involved in establishing the pattern of cell differentiation along the dorsoventral (D-V) axis of the central nervous system.

The development of the vertebrate nervous system begins with induction of the neural plate, followed by differentiation of distinct cell types at different positions along the D-V axis of the neural tube. For example, in the spinal cord, motor neurons are located ventrally, whereas commissural neurons and neural crest cells appear in dorsal positions. The distribution of each of these cell types is bilaterally symmetric with reference to the midline of the neural tube.

Cells at the midline of the neural plate cells give rise to a specialized region called the floor plate. In collaboration with Jane Dodd's laboratory, we have developed functional assays showing that the floor plate differentiation is induced by underlying mesodermal cells of the notochord. We have also found that specialized properties of the floor plate cells influence the development of other neural cells. A diffusible chemoattractant released by the floor plate orients the growth of a subset of developing spinal cord axons. In addition, the floor plate and the notochord are sources of a polarizing signal that respecifies cell pattern along the anterior-posterior (A-P) axis of the developing chick limb, mimicking the effect of the putative morphogen retinoic acid.

These findings raised the possibility that the floor plate and notochord may also be involved in controlling the patterning of cell differentiation along the D-V axis of the developing nervous system. To test this possibility, we have used antibodies directed against cell-specific antigens to determine whether the pattern of cells in the embryonic chick nervous system changes after *a*) induction of an additional floor plate at ectopic positions in the neural tube by a notochord graft or *b*) grafting of a floor plate next to the neural tube or *c*) removal of the notochord to prevent floor plate differentiation. Such manipulations cause marked changes in the fate and position of neural cell types. For example, *a* and *b* result in the appearance of ectopic motor neurons, and *c* results in the absence of motor neurons.

Our results suggest that the pattern of cell differentiation within the neural tube is established by signals originating at the floor plate and notochord. For example, the notochord and floor plate could act as local sources of a factor that diffuses through the adjacent neural epithelium, establishing a concentration gradient with its high point at the ventral midline. In this scheme, the differentiation of neural epithelial cells into distinct classes during normal development would be controlled by the signal concentration to which they are exposed.

A similar gradient model has been proposed to explain the pattern of cell differentiation along the A-P axis of the developing chick wing bud. A-P pattern in the bud appears to be under the control of a specialized region of posterior mesenchyme known as the zone of polarizing activity (ZPA), which can respecify A-P polarity when grafted to ectopic sites. Retinoic acid mimics the effects of the ZPA and appears to be distributed in a graded manner along the A-P axis of the limb bud, with its highest concentration in the posterior mesenchyme. On this basis, it has been suggested that retinoic acid functions as an endogenous morphogen involved in establishing axial polarity in the developing chick limb.

The notochord and floor plate, but not other regions of the neural tube, mimic the action of the ZPA and retinoic acid in respecifying digit pattern in the chick limb. Moreover, our biochemical studies show that the floor plate can synthesize morphogenetically active retinoids *in vitro*. The ability of the notochord and floor plate to control the pattern of cell differentiation in both the developing limb bud and the neural tube may therefore have a common molecular basis, possibly involving retinoids.

Although these studies implicate the floor plate as a source of retinoids, definite evidence that the plate, or in fact any vertebrate tissue, can release morphogenetically active retinoids is still lacking. We therefore devised a sensitive and specific assay to detect the release of morphogenetic retinoids from neural tissues. This has provided further evidence that the floor plate does indeed release retinoids. The basis of this assay is to use a fragment of DNA, which acts as a retinoic acid–responsive element, to drive expression of a reporter gene in tissue culture cells. Using F9 cells transfected with the retinoid response element placed upstream of the β-galactosidase (*lacZ*) gene, it has been possible to coculture floor plate, dorsal neural tube, and other chick and rat tissues, and to monitor retinoid release by the appearance of histologically detectable β-galactosidase in cells adjacent to the test tissue.

Rat or chick floor plate tissue was found to be a potent inducer of *lacZ* activity in nearby F9 cells. Dorsal neural tube tissue also induced *lacZ* expression, but the incidence of labeled cells, when compared with that observed with the floor plate, was markedly reduced. The intensity of *lacZ* expression was also reduced. These results show that the floor plate is considerably more effective than dorsal neural tube in inducing *lacZ* expression, and support the idea that the floor plate may be a local source of retinoids in the developing neural tube.

The evidence that the floor plate is a local source of morphogenetically active retinoids, together with the finding that it has polarizing activity within the neural tube and limb, raises the question of whether the polarizing activity of the floor plate within the neural tube is mediated by retinoids. To address this, we grafted retinoic acid–soaked beads adjacent to the neural tube of chick embryos. When bead implants were introduced into early embryos, defects in neural tube formation were frequently observed. However, the use of antibodies that reveal the pattern of cell differentiation within the neural tube indicates that a local source of retinoic acid does not mimic the effect of a floor plate graft.

There are several possible reasons for the failure of retinoic acid–impregnated beads to mimic the floor plate. First, retinoic acid may be synthesized and released by its floor plate but may not, in fact, be involved in neural tube patterning. Second, retinoic acid may act in conjunction with another floor plate–derived signaling molecule. And third, the concentration of retinoic acid diffusing from the bead may be too low to cause changes in cell pattern. We are currently investigating these possibilities.

Taken together, our findings indicate that a cascade of inductive interactions occurring early in the development of the vertebrate nervous system has a central role in the patterning of neural cells. Signals from the notochord induce the floor plate at the midline of the neural plate and neural tube. The floor plate and notochord then appear to control the differentiation of other neural cells and to contribute to the D-V pattern of the nervous system. Thus the pattern of cell differentiation in the neural tube, as in many other developing tissues, appears to depend critically on the organizing properties of specialized cell groups.

The Lactose Permease of Escherichia coli: *A Paradigm for Membrane Transport Proteins*

H. Ronald Kaback, M.D.—Investigator

Dr. Kaback is also Professor of Physiology and 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. Among his honors is the Kenneth Cole Award of the American Biophysical Society.

A critical basic biological problem that remains unsolved is the mechanism of energy transduction in biological membranes. A wide range of seemingly disparate phenomena, such as oxidative phosphorylation, bacterial motility, and solute accumulation against a concentration gradient (secondary active transport), are driven by bulk-phase, transmembrane electrochemical H^+ or Na^+ gradients. However, the molecular mechanism by which energy stored in such gradients is transduced into work or energy-rich compounds (e.g., ATP) remains unknown. In order to gain insight into this process, studies in our laboratory have focused on the lactose (*lac*) permease of *Escherichia coli* as a paradigm.

The ability of *E. coli* to accumulate the disaccharide lactose and other β-galactosides against a large concentration gradient depends on the *lac* permease, a very hydrophobic cytoplasmic membrane protein that catalyzes the coupled translocation of these sugars and H^+ with a stoichiometry of unity (i.e., symport or co-transport). Under physiological conditions, where the H^+ electrochemical gradient across the cytoplasmic membrane is interior negative and/or alkaline, *lac* permease utilizes free energy released from downhill translocation of H^+ to drive accumulated β-galactosides against a concentration gradient, the magnitude of which is directly related to that of the H^+ gradient. In the absence of an H^+ gradient, the permease catalyzes the converse reaction, utilizing free energy released from downhill translocation of β-galactosides to drive uphill translocation of H^+ and generating an H^+ electrochemical gradient, the polarity of which depends on the direction of the concentration gradient of the substrate.

Encoded by the *lacY* gene—the second structural gene in the *lac* operon—the permease has been solubilized from the membrane, purified to homogeneity, and reconstituted into proteoliposomes in a fully functional state. In addition, we have presented evidence that the permease is functional as a monomer. Circular dichroic measurements demonstrating that purified permease is about 80 percent helical, and hydropathy analysis of the deduced amino acid sequence, suggest a secondary structure in which the protein is predicted to have 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. Through other approaches, we have confirmed the general features of the model and demonstrated that the amino and carboxyl termini are on the cytoplasmic surface of the membrane. Moreover, strong exclusive support for the topological predictions of the 12-helix model is provided by Calamia and Manoil's studies on an extensive series of *lac* permease–alkaline phosphatase fusion proteins.

The topology of polytopic membrane proteins is thought to result from either the orientation of the first amino-terminal hydrophobic domain in the membrane or from topogenic determinants dispersed throughout the molecules. We have now studied the insertion and stability of in-frame deletion mutants in *lac* permease. So long as the first and last putative α-helical domains are retained, stable polypeptides are inserted into the membrane, even when an odd number of helical domains are deleted. Moreover, when an odd number of helices are deleted, the carboxyl terminus remains on the membrane's cytoplasmic surface. Thus relatively short carboxyl-terminal domains of the permease appear to contain topological information sufficient for insertion in the native orientation. Finally, permease molecules devoid of even or odd numbers of putative transmembrane helices retain a specific pathway for downhill lactose translocation.

Previous experiments indicate that amino acid residues 396–401 at the carboxyl terminus of the last putative transmembrane helix of *lac* permease are important for protection against proteolytic degradation and suggest that this region of the permease may be necessary for proper folding. Termination codons have now been substituted sequentially for amino acid codons 396 to

401 in the *lacY* gene. With respect to transport, permease molecules truncated at residue 396 or 397 are completely defective, while those truncated at 398 through 401 exhibit 15–25, 30–40, 40–45, and 70–100 percent of wild-type activity, respectively. Wild-type permease or permease truncated at position 401 is stable, while molecules truncated at residues 400 down to 396 are degraded at increasingly rapid rates. Thus either the last turn of putative helix XII or the region immediately distal is important for proper folding and protection against proteolytic degradation.

When the *lacY* gene is restricted into two approximately equal-size fragments and subcloned individually or together under separate *lac* operator/promoters, the permease is expressed in two portions: 1) the amino terminus, the first six putative transmembrane helices, and most of putative loop 7; and 2) the last six putative transmembrane helices and the carboxyl terminus. Remarkably, cells expressing both fragments transport lactose to a steady-state level of accumulation at about 30 percent of the rate of cells expressing intact permease. In contrast, cells expressing either portion of the permease independently do not transport lactose.

Since *intact* permease is completely absent from the membrane of cells expressing *lacY* fragments either individually or together, transport activity must result from an association between independently synthesized pieces of *lac* permease. If the gene fragments are expressed individually, the amino-terminal portion of the permease is observed sporadically and the carboxyl-terminal portion is not observed. When the gene fragments are expressed together, polypeptides identified as the amino- and carboxyl-terminal moieties of the permease are found in the membrane. The results indicate that the amino- or carboxyl-terminal halves of *lac* permease are proteolyzed when synthesized independently and that association between the two complementing polypeptides leads to a more stable, catalytically active complex. More recent experiments demonstrate that the permease can be split in a similar fashion, with comparable results, between putative helix I and the rest of the protein or within putative helix VII.

Notwithstanding the importance of high-resolution structure, site-directed mutagenesis can be used to delineate amino acid residues important for active transport. Over 100 single–amino acid replacements have been made in the permease, and about 70 percent of the mutations have no significant effect on permease activity. Therefore it is unlikely that individual amino acid replacements result in large conformational alterations. Arg302 (putative helix IX) and Lys319, His, and Glu325 (all putative helix X) are critically involved in lactose-coupled H^+ translocation and/or substrate binding and recognition. Moreover, molecular modeling studies suggest that the four residues may be sufficiently close to form an H-bond network.

These findings—the specific transport properties of the mutants, and construction of mutants in which the putative geometric relationship between the residues is altered—have led to the suggestion that the four residues may function in a type of H^+-relay mechanism. In contrast, Cys residues, long thought to play a central role, are not involved in substrate binding or H^+ translocation. Thus the Cys residues in the permease can be mutagenized individually or simultaneously, and significant activity is retained.

Complex 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.

THE expression of immunoglobulin genes is limited to only one cell type in the body, namely B lymphocytes. In addition to exhibiting this cell-type specificity, the expression of these genes is temporally regulated during lymphocyte development, initiating in the pre-B and B cell stages and later increasing as those cells differentiate to become mature plasma cells. Results from many laboratories have led to the conclusion that a major component of this regulation occurs at the level of transcription, the process of creating an RNA copy of the gene. Obtaining a detailed understanding of the transcriptional control of immunoglobulin genes will aid in elucidating the mechanisms of transcription regulation of other genes. Thus the study of immunoglobulin gene transcription should shed light on a variety of transcriptional control processes, including those that go awry and result in deleterious consequences (e.g., the transcriptional activation of oncogenes, leading to cancer).

Studies of the structure of immunoglobulin genes and of the DNA sequences that influence their rate of transcription in lymphocytes indicate that there are at least two major transcriptional regulatory signals (elements) within the genes. The first is the promoter, a conventional DNA element located close to the site of transcription initiation. The second is an enhancer, which stimulates activity of the promoter. Although many enhancers are found close to promoters, the enhancers within the immunoglobulin genes are located in the central portion of the genes, a few thousand bases from the promoter elements.

Since the activities of the immunoglobulin promoters and enhancers are restricted to B lymphocyte cells, it has been suggested that those cells may uniquely express transcription factors (proteins) that activate those control elements. One of our prime interests is to understand the mechanisms underlying the B cell–specific expression of the enhancer located in the immunoglobulin heavy-chain gene.

Our previous work and that of others has led to the drawing of a detailed enhancer map. This map provides the precise locations for the binding of these regulatory proteins and additional information as to how those proteins act. The enhancer is a relatively small (200 base pairs), but exceedingly complex, segment of DNA. Many of the perhaps nine or more distinct proteins that bind the enhancer are found in multiple cell types, even in cells where the enhancer is normally inactive. It is assumed that some of these proteins act to stimulate enhancer activity, while others may function to repress it.

Recently our efforts have been directed toward the cloning and functional characterization of the genes that encode these enhancer-binding proteins. Thus far we have cloned segments (i.e., cDNAs) that correspond to at least six (possibly seven) distinct genes that encode enhancer-binding proteins. We are using these gene segments to manipulate and characterize the proteins, both structurally and functionally.

Two of the encoded proteins (E2–5 and TFE3) are involved in a fascinating transcriptional regulatory scheme. In B cells the situation is relatively straightforward: both E2–5 and TFE3 bind the enhancer and act in concert to activate transcription. In non–B cells, the situation is more complex. *In vivo* experiments suggest the presence of an additional repressor protein that binds the enhancer and precludes binding of E2–5.

Binding of this putative repressor has two effects. First, the enhancer is less active due to the absence of bound E2–5 protein. Second, the repressor has the ability to attenuate, at a distance, the ability of the TFE3 protein to function. Thus the presence of the repressor in non–B cells results in the shutdown of both E2–5-mediated and TFE3-mediated activation. These effects can be overcome by artificially overproducing the E2–5 protein in non–B cells. Presumably this overexpression is sufficient to displace the bound repressor.

Because of its key role in transcriptional regulation, we are interested in identifying and characterizing the repressor. Currently, however, we have only genetic evidence that it exists. Our recent isolation of a cDNA that encodes yet another

enhancer-binding protein, ZEB, may solve that problem; we are exploring the possibility that ZEB is the repressor. It binds the same region of the enhancer that is bound by E2–5 but does not bind some of the other sites (found in other genes) that are bound by E2–5. Moreover, not all sites bound by ZEB will bind E2–5. This overlapping, yet distinct, binding specificity should allow us to evaluate whether ZEB (or something like ZEB) is the repressor.

We have demonstrated that one component of the heavy-chain enhancer's cell-type specificity results from a balance between positive- and negative-acting transcription factors. In *Drosophila*, such protein gradients are thought to establish transcriptional repertoires that dictate pattern formation during embryonic development. Hence, although our findings contradict the model initially proposed that only lymphoid cells possess the appropriate positive-acting factors, they are not without precedent. In fact, it appears that the enhancer uses both types of mechanisms to mediate lymphoid specificity. The protein Oct-2 activates the enhancer (through its binding site) and is found only in lymphoid cells. Moreover, we have recently isolated the gene for an additional B cell–restricted enhancer-binding protein. Experiments are under way to evaluate this protein's role in enhancer activity.

Although we are just beginning to scratch the surface in our understanding of transcriptional control, the immunoglobulin heavy-chain enhancer has served as an important paradigm by reflecting its potential complexity. It is likely that many other genes employ such protein gradients, perhaps combined with cell type–specific factors, to mediate their transcriptional activities. It is now thought that a group of proteins, all highly related to E2–5 (encoded by the same gene), activates transcription of not just immunoglobulin genes in B cells but of muscle-specific and pancrease-specific genes as well. These very different roles for E2–5 may reflect its ability to form heterodimers with a variety of other proteins of similar type but with distinct modes of expression. This combinatorial control undoubtedly serves to increase the overall variety of transcriptional repertoires available to the multitude of cell types within an organism.

Human Genetic Diseases

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.B., B.S. degree from the University of Hong Kong Medical School and later the D.Sc. degree. After internship and residency at Queen Mary Hospital, Hong Kong, he received 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 of 1982. He is a fellow of the Royal Society (London) and a member of the National Academy of Sciences and the Academia Sinica (Taiwan).

THE focus of our research is the molecular basis of human genetic diseases affecting the hematopoietic cells. The two diseases we have studied in detail are sickle cell anemia and thalassemia. Both are the results of 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 are defining the mutations that give rise to these defects and devising DNA analysis for their detection. In addition, we are studying the factors that control the expression of globin genes in the red cell precursors and the signals that switch from fetal to adult globin gene 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 blood 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

Currently we are refining the prenatal diagnostic tests to facilitate diagnosis of β-thalassemia in those areas of the world where this disease is an important health problem. For these tests to be used routinely, simple and nonradioactive methods are needed. The polymerase chain reaction (PCR), which amplifies specific segments of DNA many millionfold, has facilitated the design of rapid and nonradioactive tests for sickle cell anemia, α-thalassemia, and β-thalassemia.

Although prenatal diagnosis of sickle cell anemia and α-thalassemia is relatively simple because the genetic defects responsible for them are known and are readily detectable, diagnosis of β-thalassemia is more complex. Close to 100 different mutations have been found to cause the clinically important β-thalassemia syndromes. It is necessary to determine which mutations predominate in a particular region so the appropriate DNA test can be chosen, and consequently we have defined the common mutations in geographic areas where thalassemia is prevalent (Sardinia, other parts of Italy, Lebanon, and China). With this knowledge, simple and nonradioactive tests are being devised to detect the various defects in these different areas.

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 these chains is precisely coordinated during development. The embryonic ϵ- and ζ-globin chains, 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 takes over from the γ-globin chain as the predominant globin chain. The factors that control the expression of the β-globin gene in the bone marrow cells and coordinate the developmental expression of the globin genes during development have not been elucidated. We are now studying the DNA sequences and protein factors that govern the expression of these genes.

We have defined some of the sequences necessary for the tissue-specific expression of the globin genes. A DNA segment as short as 36 base pairs appears to enhance tissue-specific globin

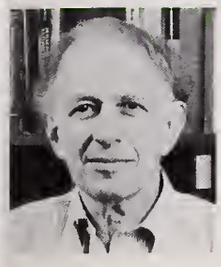
gene expression. These regions have been joined to the β-globin gene and are capable of augmenting globin gene expression in tissue culture cells and in transgenic mice. In addition, nuclear extracts from cell lines with erythroid characteristics are being examined for factors that bind to putative control regions of the globin genes. A nuclear factor from erythroid cells appears to bind to a tandem repeat of a consensus recognition sequence for a class of transacting factors called AP1. We are attempting to isolate this erythroid-specific AP1-like protein. Our long-term goal is to understand the cellular processes controlling the tissue-specific expression of the different globin genes and the signals that control the switch from fetal to adult globin. This knowledge may lead to better approaches for correcting the genetic defects in hereditary hemoglobin disorders.

Red Cell Membrane Disorders

Many hereditary hemolytic anemias are accompanied by spherocytosis or elliptocytosis. The defects in these disorders are known to lie in the membrane proteins, though the mutations were not defined until recently. To facilitate the study of these groups of hemolytic anemias, we have isolated the gene encoding for 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: one results in the absence of the mature form of protein 4.1 in the red cell, another produces an elongated 4.1 protein molecule due to duplication of three exons, and the third causes a shortened protein 4.1 due to a two-exon deletion.

We continue to define the alternate splicing mechanism by which multiple isoforms of protein 4.1 are generated. We have found at least five different regions that undergo alternate splicing. Antibodies prepared against these isoforms are being used to study their distribution and function. Some isoforms are more abundant in nucleated cells, and others, in mature red cells. Also, protein 4.1 is found in many cell types outside of the red blood cell. Thus it seems that, in addition to its role as a red cell cytoskeletal protein, protein 4.1 serves other cellular functions that need to be defined.

Cell Biological Studies of Memory



Eric R. Kandel, M.D.—Senior Investigator

Dr. Kandel is also University Professor of Physiology and Psychiatry at the Columbia University College of Physicians and Surgeons. He was born in Vienna, Austria; graduated from Harvard College, having majored in history and literature; and received his M.D. degree from New York University School of Medicine. He took postdoctoral training with Wade Marshall at the Laboratory of Neurophysiology at the NIH and with Ladislav Tauc at the Institut Morey in Paris. He was founding director of the Center for Neurobiology and Behavior at Columbia. Dr. Kandel 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 persons, places, or things. Reflexive learning refers to the acquisition of procedures and motor skills.

In the past our laboratory has focused primarily on elementary forms of reflexive learning as manifest in the gill-withdrawal reflex of the marine snail *Aplysia*. We showed that this simple reflex can be modified by both nonassociative and associative learning, giving rise to short- and long-term memory, whose duration is a function of the number of training trials.

To analyze the relationship between the short- and long-term processes for nonassociative learning (sensitization), we focused on one component of the neural circuit of this reflex—the connections between the siphon sensory neuron and the gill motor neurons. Here we found that both the short- and long-term processes involve an increase in transmitter release. Whereas the short-term process reflects enhanced transmitter release from preexisting synaptic connections due to covalent modification of preexisting proteins, the long-term process results from alterations in gene expression and the growth of synaptic connections.

What molecular mechanisms contribute to declarative forms of learning? In an attempt to compare the biochemical mechanisms underlying reflexive forms of learning with those underlying declarative learning, we have turned to the hippocampus in the mammalian brain.

Since Milner's pioneering work in the late 1950s, the hippocampus has been known to be important for aspects of long-term declarative memory storage in humans and other mammals. The hippocampus may be essential for initially storing long-term memory for days or weeks before the memory trace is consolidated elsewhere, perhaps in the cerebral cortex. In 1973 Bliss and Lømo first demonstrated that a brief, high-frequency train of action potentials in the perforant path increases the excitatory synaptic potential in the granule cells. The increase can last for hours and, under some circumstances, even for weeks. They called this facilitation *long-term potentiation*, or LTP. Later studies showed that LTP occurs at each of the three major synaptic pathways in the hippocampus.

Recent studies in the CA1 region, by Tim Bliss, Charles Stevens (HHMI, the Salk Institute), Richard Tsien, and their colleagues, have provided insights into the cellular mechanisms involved in the acquisition and maintenance of LTP. Despite important differences, this form of synaptic plasticity in the mammalian brain may bear certain similarities on the cellular level to presynaptic facilitation that accompanies sensitization in *Aplysia*. In both cases synaptic connections are strengthened through an enhancement of transmitter release, and both the short- and the long-term changes occur at the same synaptic locus.

Therefore, Tom O'Dell, Seth Grant, and I have begun to address the question, What are the molecular steps involved in LTP? It has been known for several years that calcium influx through a glutamate receptor of the *N*-methyl-D-aspartate type is critical for the induction of LTP. Less is known, however, about the subsequent biochemical steps responsible for LTP expression and maintenance. Several studies using various kinase inhibitors have implicated as important events following calcium influx the activation of two serine-threonine protein kinases: kinase C (PKC) and calcium/calmodulin-sensitive kinase II (CaMK-II). Consistent with the involvement of PKC and CaMK-II in LTP, high levels of these kinases are found in hippocampal pyramidal cells.

In addition to these serine-threonine kinases, hippocampal neurons also express high levels of protein tyrosine kinase activity. Indeed, the brain is a richer source of tyrosine kinases than any other organ, and these kinases are particularly enriched in the hippocampus. Subcellular fractionation studies showing highest levels of protein tyrosine kinase activity in crude synaptic vesicle fractions have suggested a role for tyrosine

kinases in aspects of synaptic physiology such as transmitter release. Although the importance of tyrosine kinases for synaptic transmission is not known, the above findings and the demonstrated interaction of tyrosine kinases with serine-threonine kinases in various nonneuronal signal transduction pathways led us to investigate whether protein tyrosine kinases are involved in the biochemical events underlying LTP.

A major limitation in the study of tyrosine kinases in the brain has been the lack of inhibitors of the sort that have proved useful in the study of serine-threonine kinases. Recently, several groups have developed two classes of inhibitors. One class (e.g., genistein) blocks the ATP-binding domain, and the other (tyrophostin) blocks the substrate-binding domain. These inhibitors block tyrosine kinase activity in lymphocytes and in other nonneural tissues at concentrations that have a negligible effect on serine and threonine kinases. To determine whether these inhibitors have similar specificity in the brain, O'Dell, Grant, and I examined their action on hippocampal kinase activity. We found that there, too, they inhibit tyrosine kinase activity at concentrations that have little effect on serine-threonine phosphorylation produced by kinase A, kinase C, and CaM kinase.

We next used four different inhibitors (two from each class) to examine LTP produced in the CA_1 neurons. All four blocked the induction of LTP. Moreover, their ability to block the facilitation was related to their effectiveness in blocking tyrosine kinase activity. This effect was restricted to induction; the inhibitors had no effect on the maintenance phase of LTP. Nor did they have an effect on the amplitude or duration of post-tetanic potentiation.

We also examined the specificity of the tyrosine kinase inhibitors electrophysiologically and found that several forms of synaptic plasticity attributed to other serine and threonine kinases were not affected. The enhanced firing (anti-accommodation) produced by norepinephrine (which is simulated by cAMP and thought to be mediated by kinase A) was not affected by the tyrosine kinase inhibitor genistein. Conversely, the enhanced firing (anti-accommodation) produced by muscarine (which is thought to be mediated by the diacylglycerol-IP_3 pathway) is inhibited by H-7 but not affected by genistein. Finally, the synaptic facilitation induced by phorbol esters and mediated by kinase C, which was blocked by H-7, was not affected by the inhibitors. These data suggest that the induction of LTP, or the initial steps in its maintenance, may involve a tyrosine kinase. We are now trying to explore this possibility directly by examining slices for changes in tyrosine phosphorylation after induction of LTP.

Given the known functions of tyrosine kinases in inter- and intracellular signaling processes that modulate cell growth, it is notable that the induction of hippocampal LTP, or learning at other synapses, results in growth and morphological change of neurons. With the identification of relevant kinases, their activators, and their substrates, the role of tyrosine kinases in synaptic transmission and morphological changes accompanying learning could be evaluated.

The T Cell Repertoire

John W. Kappler, Ph.D.—Investigator

Dr. Kappler is also a member of the Department of Medicine of 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. 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.

HIGHER animals are at great risk of invasion by foreign organisms. In order to avert destruction by parasites such as bacteria, they have developed complex protective mechanisms, broadly termed the immune system. The operation of the system requires, first, that it recognize an invading organism and, second, that it mount an effective response to destroy the invader.

Our laboratory is particularly interested in the first process, the means by which the immune system recognizes foreign material. It does so primarily by use of lymphocytes (white blood cells) of three different types: B cells, and T cells bearing $\alpha\beta$ and $\gamma\delta$ receptors. We have concentrated on $\alpha\beta$-bearing T cells.

Mice contain about 200 million, and humans about 1 million million such cells. All of the $\alpha\beta$ receptors on each cell are identical. From one cell to another, however, the $\alpha\beta$ receptors differ in amino acid sequence. This is because, as T cells develop, each one chooses different components from which to build its receptors.

The $\alpha\beta$ receptors are made up of five variable components: Vα, Jα, Vβ, Dβ, and Jβ. The developing T cell chooses for its receptors 1 Vα from a selection of about 50. Likewise, it chooses 1 Jα from about 50 Jα's, 1 Vβ from about 50 Vβ's, 1 Dβ from 6 Dβ's, and 1 Jβ from 13 Jβ's. Because each developing cell chooses, by chance, a different combination of Vα's, Jα's, etc., and because these components are joined together in slightly different ways by different T cells, the T cells in a given animal can express a very large number of different receptors, perhaps as many as 10^{10}.

When an invading organism enters the body, fragments of the organism (antigens) appear on cell surfaces, linked to one or more of a collection of self proteins, the products of the major histocompatibility complex (MHC). So presented, the fragments can be bound by T cells bearing the right receptors. Usually a T cell has to have just the right combination of Vα, Jα, Vβ, Dβ, and Jβ in order to bind the fragments of a particular invader. Consequently, for many such organisms, only a few T cells—perhaps 1 in 100,000—can recognize that the organism has arrived in the body. These few T cells, however, are stimulated to divide by their interaction with the invader, so in a few days they give rise to many cells able to bind fragments from the organism. Finally, T cells differentiate into cells able to destroy the invader or to stimulate other cells to do so.

A few years ago we discovered a collection of foreign materials that we named superantigens. These materials bind to MHC proteins and stimulate T cells, but unlike most antigens they will stimulate almost any T cell bearing a particular Vβ. Staphylococcal enterotoxin B, for example, will stimulate mouse T cells bearing the Vβ8's or Vβ7, or human T cells bearing Vβ3. Because there is a reasonably large percentage of T cells in mice bearing the Vβ8's and Vβ7, or in humans bearing Vβ3, these superantigens stimulate a very large percentage of T cells at once, perhaps as many as 10 percent of all cells in a given individual. This gross T cell stimulation causes overactivation of the cells, massive release of T cell mediators, and consequent illness.

A number of organisms make superantigens. Staphylococci and streptococci are among the bacteria that produce them. Superantigens are also made by mycoplasma and by mice themselves. We have been studying how these superantigens bind to MHC and to the Vβ components of the T cell receptors. Our data show that the superantigens made by mice, and by staphylococci, bind to Vβ at an exposed face of the molecule well away from the part of the T cell receptor that is thought to bind to conventional antigens plus MHC. There are also some indications that the staphylococcal toxins bind to some human MHC proteins better than others. Since human beings differ widely in the sequences of the MHC proteins they make, perhaps this is why some individuals get food poisoning when they eat staphylococcus-contaminated egg salad, for example, whereas others at the same picnic eat the same egg salad without getting sick.

We have also been studying the structures on

the bacterial toxins that cause the toxins to bind to MHC and Vβ's. So far we have identified three sites, one that seems to be concerned with binding to MHC and two that seem to be involved with Vβ binding.

Recently we identified the superantigen molecules made by mice. These turn out to be the products of viruses that, over the course of evolution, have become integrated into mouse DNA. A similar virus is transmitted from mice to their babies in milk, causing breast cancer. It is likely that T cell stimulation via the viral superantigen is an essential middle step in mouse breast cancer development. This odd interaction between viruses, T cells, and various seemingly unrelated diseases may not be unique to mice. Currently we are trying to find out whether a similar scenario applies to some viruses in humans.

Visual System Development in Drosophila

Flora N. Katz, Ph.D.—Assistant Investigator

Dr. Katz is also Assistant Professor of Biochemistry at the University of Texas Southwestern Medical Center at Dallas. After her B.A. degree in biology from Kenyon College, she did wildlife research as a Thomas J. Watson Foundation Fellow. Following her Ph.D. degree in biology with Harvey Lodish at the Massachusetts Institute of Technology, she did research at the National Biology Institute in Bogor, Java, as a Henry Luce Foundation Scholar. She then turned to research in neurobiology with Eric Kandel and James Schwartz at Columbia University and Lily and Yuh Nung Jan at the University of California, San Francisco, under the auspices of the Helen Hay Whitney Foundation and the Damon Runyon–Walter Winchell Cancer Fund.

VISION is accomplished through the integration of signals received by the eye and communicated to the brain. The neurons of the eye, or photoreceptor cells, elaborate long processes called axons that connect in specific patterns with central target areas in the brain to generate a map of the visual world. We are interested in understanding the mechanisms by which photoreceptor cells are specified and connect with their targets during development.

The eye of the fruit fly *Drosophila melanogaster* has proved to be a particularly suitable subject for studies of this nature. Unlike the mammalian eye, the fly eye is a compound structure composed of repeating units called ommatidia, which each contain eight photoreceptor cells. Six of these cells, R1–6, send axon bundles into the optic lobe (a part of the brain specialized for visual processing) that terminate in the first synaptic zone, the lamina. The central two photoreceptors, R7 and R8, send their axons past the lamina into a second zone, the medulla, where they form their connections in two distinct layers. Our efforts have focused on understanding the generation of connections by the R7 and R8 cell types.

To visualize the photoreceptor axons, we fill those originating from any selected population of ommatidia with a dye (or an enzyme that can be viewed by its reaction product). The pathway and termination zone of each R7 and R8 axon can be scored individually, and the path followed during development can be reconstructed. By visualizing the axons in this manner, those that have followed unusual routes or terminated in incorrect locations can be identified.

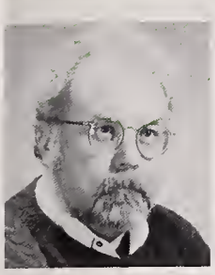
We have been studying a mutation, *nac (neurally altered carbohydrate)*, that causes a cold-sensitive disruption of the differentiation of the fly's photoreceptor cells and other selected neurons. We have shown that this mutation affects the expression of a neural-specific glycan in all neurons of the fly, although mutant flies are viable and show only selective defects in sensory behavior. When *nac* photoreceptor cells are filled with dye, the axons of R7 and R8 can be seen to enter the medulla correctly, but they show increasing defects as they approach their terminal zones. Some axons miss their targets altogether. Conversely, small groups of ectopic fibers have been seen to separate from the main axon bundles and take unorthodox routes across the medulla, nonetheless arriving in their correct terminal zones. These behaviors are likely to result from both an altered terrain and an alteration on the surface of the photoreceptor axons.

To identify additional mutations that might affect these processes more specifically, we combined our axon-filling technique with a genetic strategy. Mutations were generated by the insertion of a mobile piece of DNA, called a transposable element. When the element inserts into a gene, it inactivates it, causing a mutant phenotype.

The expression of the mutated gene can be inferred from the expression of a reporter gene carried by the transposable element, which appears to be regulated by the same controls that tell the endogenous gene where and when to be expressed. We first selected stocks of flies in which the reporter gene was expressed in the developing photoreceptor cells or in their target cells in the optic lobe. We then asked if the insertion of the mobile element had caused a mutation that affected the differentiation of the photoreceptor cells or the ability of their axons to reach the correct target sites in the brain. Several interesting mutations have been isolated in this screen.

The manner in which these mutations were generated—by the insertion of a marked piece of DNA—has furthered genetic and molecular studies now in progress. In addition, we are analyzing these mutant phenotypes with independently generated markers that highlight various portions of the axon pathways to study the rules that allow the generation of a visual map.

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 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, B.C. He is currently Secretary of the Genetics Society of America.

OUR continuing goal is to understand how the essentially one-dimensional linear information encoded in the DNA molecule is elaborated into a three-dimensional organism during the process of development. As a model system in this analysis, we are using the fruit fly *Drosophila melanogaster* and are concentrating on a group of adjacent genes in a small region of one chromosome arm. Mutations in these homeotic genes transform one organ of the fly embryo, larva, or adult into a homologous structure. For example, mutations in the *Antennapedia* gene transform the antenna of the adult fly into a leg, while mutations in the *proboscipedia* gene transform the mouthparts into legs.

A combined genetic and developmental analysis of this region of the *Drosophila* chromosome has revealed the presence of five such genes that are collectively referred to as the *Antennapedia* gene complex (ANT-C). These five genes collectively specify the proper identity of the anterior end of the fruit fly and can be viewed as developmental switches that make either/or decisions in cell fate at specific times and places in the embryo and larva.

Recombinant DNA technology has allowed the cloning of the genes and has revealed the nature of their protein products. All of the homeotic genes contain a specific DNA sequence that encodes a common protein motif called the homeobox. It has been shown that this motif acts as a DNA-binding domain and that these homeotic proteins are associated with the chromosomes in the nuclei of the developing fly. Therefore we now know that these homeotic genes act as developmental switches because their protein products serve to regulate directly the expression of other genes. What we do not fully understand is how these homeotic genes are so elegantly orchestrated in their proper spatiotemporal patterns and what is the nature of the battery of genes that are in turn regulated by the protein products of the five members of the ANT-C. Our current research is focused on these two questions.

We are using three of the five genes in the complex—*labial* (*lab*), *proboscipedia* (*pb*), and *Sex combs reduced* (*Scr*)—in our attempts to understand spatiotemporal patterning. These were chosen because of the unique properties each displayed during our initial characterization of their respective roles in *Drosophila* development.

The *labial* Gene

Using a minigene construct and P-element-mediated transformation, we have been able to ameliorate completely the embryonic defects associated with *lab* deficiency; however, adult transgenic animals show severe deformities, with thoracic structures replacing portions of the head capsule. We have now shown that this defect is caused not by a failure of *lab* activity in the developing adult head but by the ectopic expression of the minigene in the head anlagen, which is also associated with a failure to express the *Deformed* (*Dfd*) and *Scr* loci in their normal pattern in this same tissue. Thus it appears that *lab* can act as a transregulator of these other homeotic members of the ANT-C. This is in marked contrast to our observation that cross-regulatory interactions do not take place among any of the ANT-C homeotic loci in the embryo. It would appear therefore that the regulatory hierarchy in the embryo and adult stages is different.

The fact that the minigene shows ectopic expression only in a lab^- background points up another regulatory phenomenon. Since the resident *lab* gene is able to prevent ectopic expression, the native protein appears to be able to influence the expression of the minigene negatively: i.e., *lab* shows autogenous negative regulation. The product of the minigene, however, does not perform this function. In the construction of the minigene we used a cDNA fragment that did not include a minor alternate splice form. This alternate RNA product would produce a protein product that is six amino acids longer than the single protein encoded by the minigene. The resident gene can, of course, make both proteins. These observations lead to the exciting possibility that the longer protein product is responsible for the negative regulation and that the short form is inca-

pable of this activity. We are currently testing this hypothesis in transgenic animals.

The *proboscipedia* Gene

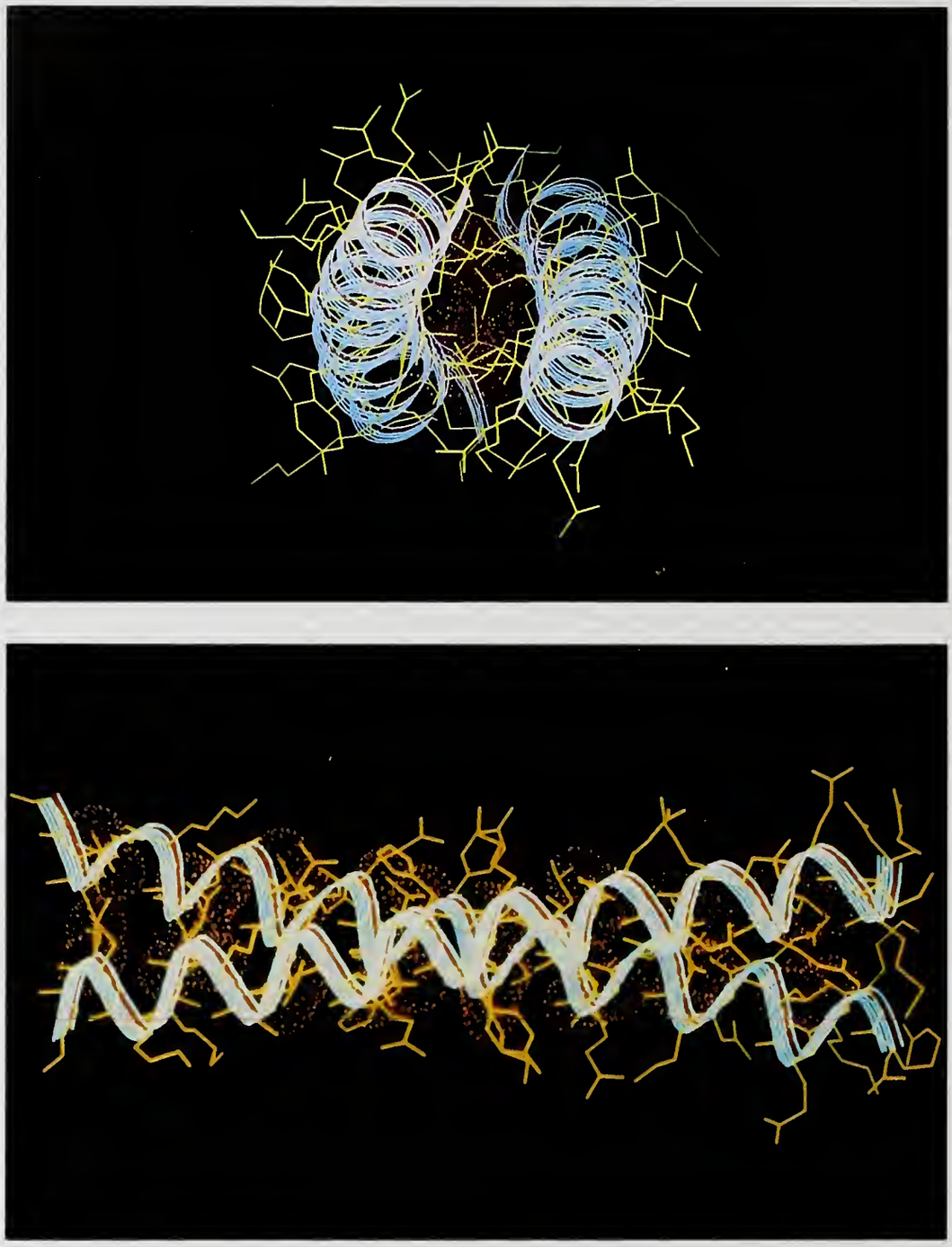
As is the case for the *lab* locus, we have constructed a minigene of the *pb* transcription unit, returned it to flies, and affected a rescue of the adult mouthpart-to-leg transformation. However, unlike *lab* we were unable to identify sequences upstream of the start of transcription that were sufficient to drive expression of a β-galactosidase reporter gene. We have now found the fragments necessary for a *pb* pattern of expression in the major intron of the gene. Specifically, a 2-kb fragment from the intron in combination with 7 kb of upstream sequence is sufficient to drive reporter expression in a *pb* pattern. The intronic regulatory elements are evolutionarily highly conserved, and identical DNA sequences can be found in the same position in the *pb* loci of two other distantly related *Drosophila* species. We are currently defining the extent of the *pb* promoter through a deletion analysis of the appropriate sequences in both the minigene and reporter constructs. Moreover, we are testing the ability of the homologous sequences from the two related flies to drive *pb* expression in *D. melanogaster*.

The *Sex combs reduced* Gene

We have shown that *Scr* locus expression in the adult is regulated by a transvection-like mechanism. Normally *Scr* protein is accumulated in the labial and first thoracic imaginal discs. Transvection mutants also express the gene in the second and third thoracic segments. This deregulation occurs by a misexpression of the copy of the gene held in trans to the chromosome that bears the mutant allele. To locate the sequences that respond to the transvection signal, we have cloned portions of the *Scr* regulatory DNA into appropriate reporter constructs. These are now in flies, and we are in the process of determining their expression patterns in both normal and transvection-inducing mutant backgrounds.

Regulated Ectopic Expression of the ANT-C Homeotics

We have built constructs in which each of the ANT-C homeotic loci are driven by the hsp70 promoter and have determined the effects of global expression of each of these genes on morphology and the expression of the other homeotic loci. Now we are using a two-element system developed in the laboratory of Norbert Perrimon (HHMI, Harvard Medical School) to fine-tune this approach. The promoter elements of *lab* and *pb* have been cloned in front of the yeast GAL4 structural gene. Additionally cDNAs and genomic fragments from all five ANT-C homeotics have been cloned downstream of a promoter that is regulated by the GAL4 protein. These constructs will be transformed separately into flies. It will then be possible to express each of the homeotic structural genes in the domains of the *lab* and *pb* promoters by crossing flies containing either of the promoter-GAL4 constructs to animals carrying each of the receptor-homeotic chimeras. These constructs and crosses will allow us to determine if the precise ectopic expression of one homeotic gene has any effect (interference or amelioration) on the locus normally expressed in that spatiotemporal environment.



Images (top and side views) of the x-ray crystal structure of a peptide corresponding to the leucine zipper of the yeast transcriptional activator GCN4, determined at 1.8 Å resolution, in collaboration with Thomas Alber's research group at the University of Utah School of Medicine. The leucine zipper folds as a parallel coiled coil, with two helices supercoiled around each other. The crossing angle of the helices is about 20°.

Research of Peter S. Kim.

Protein Folding and Macromolecular Recognition

Peter S. Kim, Ph.D.—Assistant Investigator

Dr. Kim is also Associate Member of the Whitehead Institute for Biomedical Research, Assistant 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.

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. In particular, nuclear magnetic resonance (NMR) assignments have been made for essentially every proton in BPTI by Kurt Wüthrich and co-workers (Eidgenossische Technische Hochschule, Switzerland), and several high-resolution crystal structures of the protein are available. In addition, Thomas Creighton (European Molecular Biology Organization, Germany) has characterized the folding of BPTI in terms of disulfide bond formation.

It is difficult to determine the structures of protein folding intermediates because protein folding is a cooperative process. Indeed, a high-resolution structure has not yet been determined for any protein folding intermediate. 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 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. It appears that much of the folding pathway of BPTI can be explained by the formation of native-like subdomains in these two intermediates. If our results are general, and native-like subdomains turn out to be key determinants of protein folding, then solving the protein folding problem might be reduced in large part to identifying and understanding subdomains of native proteins.

In addition to peptide models, we are studying recombinant BPTI molecules that contain a subset of the native cysteine residues. We are testing our understanding of the folding pathway by making mutant BPTI molecules that should alter it and are developing new methodologies for studying it. The overall goal is to learn in detail how this small protein folds, in structural, thermodynamic, and kinetic terms.

Theoretical attempts to model the protein folding process from first principles are severely hindered because of the enormous number of interactions to consider and because the calculations must necessarily be very accurate (the stability of most proteins is determined by a tiny difference between large energies favoring and opposing folding). One of the major complications is solvent: interactions between water and protein molecules, and between water molecules themselves, are numerous and difficult to calculate accurately. We are therefore trying to study peptides in the gas phase. Using a relatively new method called laser desorption, we are putting peptides into the gas phase for study by optical measurements—e.g., fluorescence and circular dichroism. Experimental studies of the structure

and dynamics of short peptides in isolated, well-defined conditions such as the gas phase will provide a valuable bridge to theoretical simulations.

Macromolecular Recognition

In this area, we have focused on the leucine zipper class of DNA-binding transcriptional activator proteins, originally identified by Steven McKnight and co-workers (HHMI, the 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. Recent x-ray crystallographic studies of this peptide (with Tom Alber's group, University of Utah) 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. We are using NMR to investigate the dynamics of the leucine zipper dimer by measuring amide proton exchange rates. Combinatorial mutagenesis is being used (with Robert Sauer's group, Massachusetts Institute of Technology) to investigate the sequence requirements for dimerization by this leucine zipper.

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. We are investigating the mechanism and structural basis of this specificity in detail.

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.

RNA 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 genetic and biochemical implications of this fact for a virus, and in any mechanistic similarities or differences in genetic processes between organisms with RNA and DNA genomes. In addition, we are exploring the mechanisms of RNA packaging, RNA replication, and RNA recombination in the genome of poliovirus and other viruses. We are also interested in the interactions between viruses and their host cells, especially in the area of RNA-protein biochemistry.

Many of our genetic studies utilize poliovirus, a small icosahedral virus with an RNA genome only 7,500 nucleotides in length. We have shown, for example, that RNA recombination among poliovirus genomes occurs at sufficient frequency that 1 out of every 25 is a recombinant. In contrast to the breaking and joining of preexisting molecules that leads to DNA recombination, RNA recombination occurs during RNA synthesis. This results from the switching of parental template RNAs by RNA polymerase. We are developing an *in vitro* RNA recombination system with a view to asking more detailed mechanistic questions. For example, can the RNA replication proteins switch RNA strands themselves or are other proteins required? How do the RNA strands involved in a recombination event find each other and align in the proper way? Can RNA polymerases from RNA viruses other than poliovirus also accomplish recombination?

Further investigation into the incidence of RNA recombination among other RNA genomes besides that of poliovirus will be facilitated by the use of physical rather than genetic assays. We are using the polynucleotide chain reaction to screen the progeny of crosses of various RNA viruses, phage, and subviral parasites that are not amenable to direct genetic analysis. We hope to increase 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, Jim 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 virus RNA into the final virion structure, nor do we know the structural requirements of the RNAs and proteins participating 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 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 intact empty particles and smaller subviral particles from both mutant and wild-type poliovirus-infected cells. We hope to find a correlation between the mutant defects in RNA packaging and a defect in the binding of one of these particles to RNA *in vitro*. Such a correlation would certainly strengthen the idea that the particle is an intermediate in viral assembly, possibly directly responsible for packaging the viral RNA.

We are also using *in vitro* RNA-binding assays to examine the sequence specificity of the viral RNA binding to the subviral protein particles that are candidate intermediates in viral assembly, to understand any RNA structure or sequence specificity of viral packaging.

To investigate the role of host cells in the propagation of the genomes of RNA parasites, we are extending the study of RNA genetics to yeast. 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. Virus-like particles containing double-stranded RNA genomes replicate in the cytoplasm of yeast cells in large numbers. Their genetic analysis, however, will depend on our ability to make defined mutations in their RNA genomes. To this end, we are constructing perfect DNA replicas of these RNA genomes.

It has been necessary to develop the technology of introducing RNA directly into yeast cells. Using RNA molecules encoding the luciferase protein of fireflies, we have been able to demonstrate the direct introduction of RNA into yeast, and to determine how best to transfect RNA viral genomes into yeast cells. By introducing into cells mutant RNA viral genomes, we will be able to study any mutation-induced defects in the resulting infection. And, by studying the effects of yeast mutants, we will be able to assess contributions of the host cell to the infectious process.

Signal Transduction by Adrenergic Receptors

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 positions 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 eight types of adrenergic receptors have been cloned. These receptors have different functional properties and play different roles in the sympathetic nervous system.

All of these receptors, however, are structurally similar, having seven hydrophobic domains that are thought to be membrane spanning. Mutagenesis studies have revealed that the hydrophobic domains are involved in forming the catecholamine-binding pocket. The cytoplasmic domains, which are hydrophilic, are involved in activating G proteins and in desensitizing the receptors.

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. Mutagenesis studies are identifying amino acid residues involved in binding subtype-specific ligands. These studies will help to define the boundaries of the ligand-binding pocket and possibly provide insight into the differences between agonist and antagonist binding.

A long-range goal in my laboratory is to characterize the three-dimensional structure of the β_2-adrenergic receptor and to understand how the structure changes during signal transduction. Our current efforts are focused on developing methods to produce large quantities of pure, functional receptor protein for biochemical and biophysical studies. We are attempting to increase production of β_2 receptor and to improve the purification procedure.

Efforts include designing expression vectors and determining the optimum conditions for growth of tissue culture cells. Efforts at improving the efficiency of purification involve developing new ligand affinity resins and making recombinant receptors with minor structural modifications that facilitate purification.

Receptor Biosynthesis

The primary amino acid sequence of a receptor contains all of 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 receptor protein.

We have developed a cell-free expression system capable of synthesizing functional β_2-adrenergic receptor. This system was used to study the process by which the β_2 receptor is inserted into the endoplasmic reticulum and folded into a functional protein. Research thus far has shown that the receptor is nonfunctional immediately after translation and translocation into the endoplasmic reticulum. To produce a functional receptor, additional processing is needed. ATP, intact microsomal membranes, and a high-molecular-weight cytosolic factor are required for this processing. We are attempting to determine the nature of this processing by identifying 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. A great deal has been learned about the role here of phosphorylation, but the mechanisms for receptor removal are poorly understood. We have obtained antibodies to the β_2-adrenergic receptor and are using them to characterize the process of agonist-mediated receptor internalization. We hope to learn about the molecular mechanism by which these agonist-occupied receptors are identified and targeted, either for reversible sequestration into an endosomal compartment or for destruction.

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 the 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.

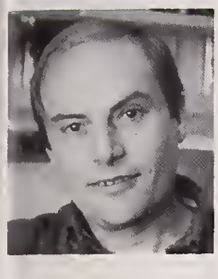
As a prototype, we have cloned the juncture between chromosomes 14 and 18 that is present in the most frequent form of human lymphoma, follicular-type B cell lymphoma. This translocation occurs early in the development of a B cell and introduces a newly discovered gene, *Bcl-2*, into the Ig locus. A 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 B 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 *Tcl-3* from chromosome 10, are not normally expressed in T cells. Instead their function is diverted from their normal sites to T cells. This provides an important opportunity to unravel how the redirection of these regulatory genes into another cell type 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. In characterizing such defects, the challenge is to develop technologies that analyze large expanses of the human genome. 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 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.

Molecular Genetics of Neuromuscular Disease



Louis M. Kunkel, Ph.D.—Associate 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.

OVER the years our laboratory has worked on how abnormalities of the protein dystrophin result in the clinical symptoms of Duchenne and Becker muscular dystrophy. From our initial identification of the dystrophin gene, we have turned to the normal protein function and to how this might be restored to mitigate disease symptoms. Our efforts have centered on proteins related to dystrophin by structure or function. Not only may these replace absent dystrophin in diseased muscle, but they are themselves prime candidates for disease-producing alteration.

Over the past year we cloned two dystrophin relatives and studied their function in diseased and normal persons. We also designed more accurate ways of detecting mutations within the dystrophin gene and pinpointed dystrophin in the central nervous system. We have also hypothesized a role abnormal dystrophin might play in cognitive function.

After dystrophin had been cloned and sequenced, it was seen to belong to a family of proteins with cytoskeletal function. A group at Oxford University, using clones for dystrophin, was able to isolate a part of a member of the family. This protein, encoded by a locus on chromosome 6, was partly sequenced and shown to be 85 percent identical to dystrophin.

Using the published sequence, we cloned a part of this locus and expressed the protein in bacterial cells. Antibodies directed against the new protein were produced. Used as a probe against normal muscle, the antibodies detected a protein the same size as dystrophin, and it was also found in the muscle of Duchenne patients. We called it dystrophin-related protein, or DRP.

Using the antibodies to probe for DRP, we were able to show that many different tissues expressed it and that the highest levels were in developing fetal muscle. By microscopic localization, we showed further that DRP colocalized in muscle with the acetylcholine receptor. In a developing muscle the protein colocalizes with dystrophin, but as the muscle matures, DRP migrates along with the acetylcholine receptor to the neuromuscular junction.

We are currently trying to unravel why a protein that is so similar to dystrophin has such a specialized function in muscle and how we might interfere with normal function to make DRP work more like dystrophin. We are also searching for a neuromuscular disease that might be caused by abnormalities of DRP.

Our antibodies directed against dystrophin detected a second protein, which we have identified and partly cloned. This large molecule was found exclusively in the brain. Using our dystrophin antibodies as a probe, we were able to clone pieces of cDNA that encoded this protein. As part of our strategy in characterizing new members of the dystrophin family, we mapped the encoding locus to chromosome 5. A search of the human linkage map revealed that the gene altered to cause spinal muscular atrophy has been localized to chromosome 5q12. We have initiated a collaborative effort with David Ward (Yale University) to localize further our dystrophin-related protein on chromosome 5. We are also attempting to generate highly polymorphic markers surrounding the locus to map this gene relative to SMA mutations. This latter effort will be in collaboration with Conrad Gilliam (Columbia University), who was involved in the original mapping of SMA to chromosome 5.

One of the curious features of some Duchenne patients is mental impairment. Because virtually all patients have almost no dystrophin, it was difficult to explain the impairment in some. We showed that dystrophin was expressed in the brain, but under different regulatory control from that found in muscle. In the course of developing more efficient dystrophin-detecting antibodies, we found one that was highly sensitive for dystrophin and showed little cross-reaction with other brain proteins.

We used this antibody to localize dystrophin in the brain, both with the light microscope and the electron microscope. Unlike results in muscle, where all types of muscle reveal dystrophin, it was only found in a subset of neurons of the cerebellum and cerebral cortex. We hypothesize that an absence of dystrophin in the cortical neurons

might yield a mild mental impairment, explaining the mental phenotype observed in Duchenne patients.

Abnormalities of dystrophin are easily detected at the protein level, and nearly 70 percent of mutations that cause these abnormalities have been shown to be deletions or duplications of some part of this extremely large locus. The remaining mutations are probably single-base changes that either disrupt translation of a functional protein or affect the processing of the large primary dystrophin transcript.

We have designed primers from dystrophin's nucleotide sequence to allow polymerase chain reaction (PCR) amplification of specific regions of the protein's transcript. By looking at the transcript, we should be able to determine the nature of these other mutations and to follow them in families at risk of having inherited such mutations. This ability will improve the diagnosis of Duchenne dystrophy and complement existing methods.

Our aim for the future year is to build on work already in progress. We will continue to identify, clone, and study new dystrophin relatives with an eye toward their possible role in mitigating the effects of abnormal dystrophin. We will also study any possible role these proteins might play in other neuromuscular diseases. As more disease loci are mapped to regions of human chromosomes, our candidate gene approach should help in the rapid characterization of these diseases.

X-ray Diffraction and Computer Simulation Studies of Protein Function

John Kuriyan, Ph.D.—Assistant Investigator

Dr. Kuriyan is also Assistant Professor of Molecular Biophysics at the Rockefeller University. He graduated from Juniata College in Pennsylvania, with a B.S. degree in chemistry. He received his Ph.D. degree from the Massachusetts Institute of Technology, working jointly with Gregory Petsko and Martin Karplus at Harvard University on the dynamics of proteins. He continued in the Karplus laboratory as a postdoctoral fellow before moving to Rockefeller as a University Fellow. He is also a Pew Scholar in the Biomedical Sciences.

OUR interests are in characterizing the structures of proteins, using crystallography—namely, x-ray diffraction experiments—and computer simulations. Crystallography and computer simulations are complementary approaches to understanding protein function. The x-ray experiments provide three-dimensional structures, which are currently impossible to generate from theoretical considerations alone, and the simulations allow us to visualize the effects of extrapolations that are not experimentally feasible. We apply the knowledge gained from these studies toward the design of mutations and inhibitors to modify protein activity.

Part of our work involves applications of the molecular dynamics (MD) method to problems in protein crystallography. This powerful tool for simulating protein structure and structural relationships involves computing the forces between all the atoms in a protein and generating trajectories of the atomic motion in response to these forces. Reasonably realistic MD simulations of solvated enzymes or proteins in crystal lattices are now possible. The simulations, however, rely on rather approximate interatomic potential functions and, even with the fastest computers available today, are limited to very short timescales. We are carefully testing and evaluating the results of MD calculations through simulation of small proteins, such as crambin and ribonuclease, in the crystal environment, thus permitting detailed comparison with high-resolution x-ray diffraction data.

In related work, we are developing methods for the better treatment of dynamics in refining crystallographic structure.

Our projects in the determination of crystallographic structure encompass two major areas: one involves redox proteins and the transcriptional response to oxidative stress (in collaboration with Peter Model and Anthony Cerami at the Rockefeller University and with Gisela Storz at the National Institutes of Health), and the other is concerned with transcription factors in *Drosophila* development (in collaboration with Claude Desplan, HHMI, the Rockefeller University). We focus here on just the first topic.

One of the efforts in this area is aimed at developing an understanding of the catalytic and substrate recognition mechanisms of two related redox enzymes, thioredoxin reductase and trypanothione reductase. These are members of a widely distributed family that channel the reductive power of NAD(P)H via a protein-bound flavin (FAD), to an active-site disulfide bond. Other well-known members include glutathione reductase (for which a high-resolution x-ray structure was first determined), lipoamide dehydrogenase, and mercuric ion reductase.

Thioredoxin is a small redox-active protein that is the reductant of ribonucleotide reductase in the DNA synthesis pathway. It has diverse other functions, ranging from the light-activated regulation of enzyme pathways in plants to the catalysis of protein disulfide isomerization. In *Escherichia coli*, thioredoxin is maintained in the reduced state by the action of NADPH-dependent thioredoxin reductase (TR). Previous x-ray diffraction studies on the related enzyme glutathione reductase (GR) revealed a dimeric structure (M_r = 2 × 52,400) with four domains within each molecule: the FAD- and NADPH-binding domains, the "central" domain that is also responsible for binding FAD, and the carboxyl-terminal domain that provides the dimer interface as well as residues that are critical for substrate binding and catalytic activity.

Although TR and GR both catalyze the same chemical reaction, that of disulfide reduction in their substrates, TR is significantly smaller (M_r = 2 × 34,500) and lacks the interface domain that in GR forms part of the active site. Another fundamental difference in their sequences involves the redox-active disulfide, which in GR is part of a highly conserved hexapeptide (Cys-Val-Asn-Val-Gly-Cys) in the FAD domain, and in TR is a shorter segment (Cys-Ala-Thr-Cys) found instead in the NADPH domain.

The three-dimensional structure of TR, determined at 3.0 Å resolution by multiple isomorphous replacement (in collaboration with Peter Model and with Charles Williams at the University of Michigan) and initially refined at a resolu-

tion of 2 Å using synchrotron x-ray data, shows that despite these differences the tertiary structure of TR is closely similar to that of GR. However, the quaternary structures and active-site architectures are unrelated, and the two enzymes appear to have acquired the disulfide reductase activity independently. These results provide a striking example of modular evolution of three-dimensional protein structure and enzyme function.

Trypanothione reductase (TrypR) represents an enzyme target for drug intervention in African trypanosomiasis, Chagas' disease, and leishmaniasis. The protozoan parasites do not possess GR, and instead use a glutathione-based peptide, trypanothione, for the essential reduction of glutathione. Trypanothione is in turn reduced by TrypR, which was first characterized from an insect trypanosomatid, *Crithidia fasciculata*. This analysis revealed that it is very similar to GR in size, catalytic mechanism, and active-site structure.

The two enzymes, however, are mutually exclusive for substrate. This, combined with the known susceptibility of trypanosomatids to oxidative stress in the absence of reduced glutathione, makes TrypR a promising target for the development of therapeutic agents that are not toxic to the host cell. We have obtained nicely diffracting single crystals of trypanothione reductase from *C. fasciculata* and have recently solved the structure to 2.4 Å resolution by molecular replacement, using the structure of human erythrocyte GR as a search model (in collaboration with Anthony Cerami). We are in the process of identifying the residues that are responsible for the 10^4 enhancement in the turnover of glutathione for trypanothione in the parasite enzyme.

An extremely interesting and well-characterized redox-based regulatory system is the bacterial response to oxidative stress. Exposure of *E. coli* or *Salmonella typhimurium* to low levels of hydrogen peroxide results in subsequent resistance to high levels. This is due to the induction of 30 proteins. The OxyR protein is responsible for the peroxide induction of nine of these, including GR, catalase, and an alkyl hydroperoxide reductase that is similar in sequence to TR.

Unlike many bacterial regulatory systems, where distinct proteins are involved in sensing the environmental change and in activating transcription, OxyR is both the sensor and the transcriptional activator. The levels of OxyR do not change significantly when the cell is challenged with peroxide. Rather, OxyR undergoes rapid and reversible changes in its DNA-binding properties. Depending on which site it is bound to, OxyR acts as either an activator or a repressor of transcription. We have entered into a collaboration with Gisela Storz to crystallize OxyR and determine its three-dimensional structure, its mode of binding to DNA, and the mechanism of its response to oxidative stress. Since OxyR is member of a large family of bacterial regulatory proteins of as yet unknown structure or mechanism, our studies are likely to have broad implications.

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. (cell biology) degrees from Albert Einstein College of Medicine, did his internship and residency in pediatrics at the University of Pittsburgh, and held a fellowship in medical genetics at the University of Washington. After seven years as a faculty member at the Children's Hospital of Harvard Medical School in Boston, he joined the University of Michigan Medical Center.

WE are interested in the two major aspects of Down syndrome, namely the etiology and pathogenesis of this disorder.

Etiology

Why does a woman's risk of having an offspring with Down syndrome increase dramatically with age? There are two competing theories. The "older egg" model states that as women age, they produce more abnormal (aneusomic) eggs. The "relaxed selection" model states that as women age, they lose the ability to abort abnormal conceptuses. To determine which of these hypotheses, or whether a combination of the two, is correct, it is necessary to study many cases of Down syndrome.

Our laboratory examines the molecular aspects of such a study. We utilize polymerase chain reaction (PCR) techniques that analyze polymorphic variability due to differences in the quantity of interspersed d(A,C):d(G,T) sequences and to differences in the length and/or sequence of the 3′ end of human interspersed highly repetitive *Alu* sequences. Two major advantages of these techniques are the extensive variability of these sequences and the requirement of only small amounts of subject DNA for the PCR-based analysis. The latter point is particularly important, as blood from even small children becomes almost an inexhaustible resource with this technology.

To detect useful variability on chromosome 21, we hybridized cosmids containing flow-sorted DNA enriched for this chromosome (courtesy of Lawrence Livermore Laboratories and Imperial Cancer Research Fund) with a poly d(A,C):d(G,T) probe. In addition, we isolated several cosmids and yeast artificial chromosomes (YACs) that contain variable oligo d(A,C):d(G,T) and *Alu* sequences near the centromere of 21q. Variability detected by oligomers that abut these sequences on 21q along with variability detected by more standard restriction fragment length polymorphisms has resulted in two important (and unexpected) findings: 1) 94 percent of nondisjunction errors are maternal in origin, a significantly higher percentage than the 75 percent figure based on subjective cytogenetic observations that must have been in error; 2) a significant plurality of Down syndrome cases show no detectable crossing over on chromosome 21 in the parent responsible for nondisjunction.

It will be especially important to distinguish between meiosis I and II nondisjunction errors. To accomplish this, we have performed locus expansion on D21S16 and D21S120 and isolated various YACs that in aggregate comprise the most proximal known markers on 21q. We are currently extracting variable markers from these clones. The establishment of molecular variability near the centromere of chromosome 21 will enable us to determine whether an error is maternal meiosis I, maternal meiosis II, paternal meiosis I, or paternal meiosis II. By comparing these rates among women of different age groups, it will be possible to determine whether the older egg or relaxed selection model (or a combination of the two) is correct. Furthermore, it is too early to determine whether the lack of crossing over observed in a plurality of cases reflects no chiasmata or terminalization of chiasmata and whether this depression of crossing over is associated with advanced maternal age. To answer these questions, it will be necessary to examine variable molecular probes that span the long arm of chromosome 21, including markers at the distal tip of 21q. Whatever the results, the biology of nondisjunction is of even greater interest and complexity than imagined originally.

Pathogenesis

To isolate genes encoded by chromosome 21, we use a novel recombination-based methodology. Genomic fragments isolated from chromosome 21 are cloned into a plasmid vector with the genetic marker *sup*F. RNA isolated from a variety of human tissues was used to construct bacteriophage λ cDNA libraries, which are then infected into cells harboring *sup*F plasmids carrying individual nonrepeated genomic fragments on chromosome 21. If any member of the genic cDNA library shares homology with the ge-

nomic DNA fragment, then recombination mediated by that homology will ensue.

Following recombination between the plasmid and the bacteriophage, 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 also designed to stand alone or to interdigitate with the genomic initiative as it proceeds. In the latter case, as sequencing detects open reading frames, the recombination-based assay is designed to delineate the tissue and timing of transcription quickly and accurately and isolate the transcribed sequence.

This methodology has worked in a model experiment and resulted in the isolation of at least one gene on chromosome 21. However, in performing these studies, two arcane problems were noted.

1. Once recombination indicating transcription has occurred, further recombination between the λ phage of the cDNA library and the λ lysogen used originally to insert a single copy of the P1 *ban* gene results in frequent scrambling of the λ phage, carrying the cDNA into an unrecognizable state. This can be circumvented by cloning a copy of the P1 *ban* gene in another way than as a λ lysogen. The final successful strategy (which we accomplished recently) was to construct a bacterial strain (DK43) carrying a low-copy P1 lysogen with a functioning *ban* gene and the P1 restriction system inactivated insertionally. This strain can now be used to isolate genes on chromosome 21.

2. The bulk of cDNA (gene) libraries, made in other laboratories for other purposes, are contaminated with small amounts of ubiquitous plasmid pBR322 DNA sequences. These sequences prevent the screening of such libraries with genomic clones inserted in ColE1 (pBR322)-based plasmids by recombination. To circumvent this problem, we are cloning the *sup*F gene into an R6K-derived plasmid that lacks homology with pBR322. Since the R6K replicon is used rarely and need not be present in the strains used for propagation of λgt11 libraries, contamination of cDNA libraries with this plasmid will not occur. This will allow the screening of a wide variety of extant cDNA libraries for transcription by chromosome-specific elements.

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. 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 upon 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. It 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 and change the biological properties and pathogenicity of the virus as a result. This means that the enzyme catalyzing coronaviral RNA synthesis is also unusual, which is, indeed, suggested from the sequence of the gene encoding the enzyme. Our laboratory is investigating this novel RNA synthesis mechanism.

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.

Furthermore, we demonstrated that recombinant viruses could become a predominant virus population under certain conditions, replacing the parental viruses by a simple process of natural selection. 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.

This genetic phenomenon also has an important implication in vaccine development for diseases such as AIDS (acquired immune deficiency syndrome), since genetic exchanges may lead to genetic instability of 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 disease.

Another virus we are studying is hepatitis delta virus (HDV), a human hepatitis virus commonly associated with a severe form of hepatitis. 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 in order to infect liver cells. The 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 17,000 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, which cause a variety of plant diseases. Indeed, the similarity be-

tween 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 shown that HDV RNA represents a new class of ribozyme, which is distinct from any others known.

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 HDV RNA synthesis. We have been studying this protein's properties and functions and have shown that it interacts with HDV RNA in a specific way. HDAg may be the reason the virus causes diseases.

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 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 must make their own enzymes, since normal cells do not appear to have this type.

HDV thus provides a new perspective 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.

Human Papillomaviruses Types 16 and 18

Laimonis A. Laimins, Ph.D.—Assistant Investigator

Dr. Laimins is also Assistant 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 of the NIH.

MY laboratory is studying the molecular biology of human papillomaviruses types 16 and 18 and their relationship to neoplastic disease. HPV-16 and -18 are the probable causative agents of the many cancers of the urogenital region, in particular those of the cervix of the uterus. Infection by HPVs is, however, not sufficient by itself to induce cervical cancer; other cellular events, such as the activation of oncogenes, may also be required. We are currently analyzing the molecular mechanisms by which HPVs contribute to the development of malignancy.

Human papillomaviruses are classified into various subtypes, depending on sequence homology. If an isolate shares less than 50 percent DNA homology with other subtypes, it is classified as a new viral type. More than 66 distinct HPV subtypes have been identified. The majority of papillomaviruses (numbered according to chronological order of discovery) cause benign skin warts; only a limited subset is associated with cervical cancer (HPV-16, -18, -31, and -33). HPV-6 is responsible for many genital warts; HPV-16 and -18 are found in only a small proportion of benign lesions. In contrast, more than 95 percent of all malignant lesions contain HPV-16, -18, and -31 sequences. An important question that our laboratory is studying is what factors differentiate a virus that causes benign lesions from one capable of inducing malignancies. Although the number of infected 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.

HPV-16 and -18 have a latent period between the time of infection and the appearance of a neoplastic state. In the benign state the virus is usually found as an episome; in the malignant state the genome is usually found integrated into the host chromosome. One model suggests that this integration process is important in the development of the cancerous state, perhaps through the removal of a dominant inhibitor of transformation. Thus viral integration into the chromosome may be the first step of a multistep process in the development of neoplasia.

We have been studying how HPV viral genes contribute to the transformation process. Using *in vitro* protocols, we have transformed both human and mouse cells in tissue culture. In one set of experiments, we introduced HPV DNA into the immortalized rodent cell line NIH 3T3 and found that these cells could induce tumors after injection into mice. Since NIH 3T3 cells by themselves do not induce tumors, this demonstrates the tumorigenic potential of HPVs.

We have also investigated the ability of HPV to transform primary human epithelial cells, the natural host cell for HPV infection. HPV can immortalize these cells, which normally only have a limited life span *in vitro*. In addition, using a system that mimics the differentiation properties of epithelial cells *in vitro*, we have shown that HPV sequences alter the ability of epithelial cells to differentiate. Although cells transfected by HPV-16 and -18 continue to stratify, the expression of viral genes disrupts the normal differentiation pattern.

The morphological changes that we observe in this tissue culture system are 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. Two HPV genes, E6 and E7, seem to be required for this transformation process. This system may be an *in vitro* model for the development of cervical cancer; we hope to utilize it to identify important factors involved in controlling the development of malignancy. These observations support the etiological role of HPV in the development of cervical cancer.

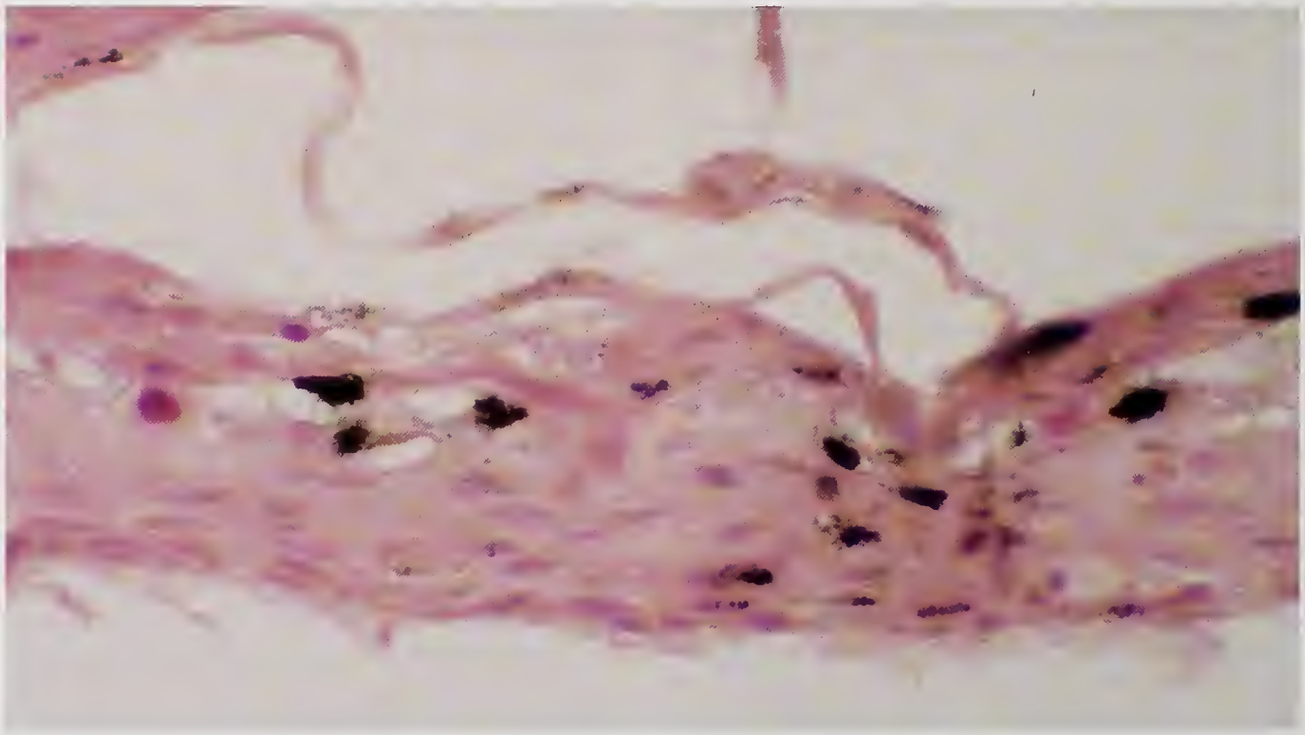
Although it is still not possible to propagate HPVs in tissue culture, we have made significant progress in this area. We have duplicated two features of a productive infection *in vitro*: the differentiation-specific amplification of viral genomes and the expression of late transcripts. A block still remains at the level of synthesis of capsid proteins, but we hope to overcome this obstacle in the near future.

To understand the life cycle of papillomaviruses, we are examining the factors involved in the tissue-specific regulation of viral gene ex-

pression and replication. We have identified HPV-18 enhancer sequences that are responsible for expression in maturing keratinocytes. The function of one of these sequences depends on the action of a cellular gene found only in epithelial cells, the keratinocyte-stimulating factor (KRF-1). We hope to understand the tissue spectrum of expression of papillomaviruses by studying the epithelial-specific mechanisms for regulation of transcription.

In collaboration with Robert Hammer (HHMI, University of Texas Southwestern Medical Center at Dallas), we have developed an animal model for papillomavirus-induced disease. Transgenic mice have been constructed that express only E6-E7 genes of HPV-18. These mice develop tumors of the urogenital tract: males develop seminal vesicle tumors; females develop what appear to be tumors of the cervix. This demonstrates that HPVs are capable of inducing urogenital tumors in animals. This model may thus be useful in the study of virally induced neoplasias.

The study of HPV will provide information on mechanisms of transformation and tumor progression, as well as tissue-specific expression and viral replication.



In situ *hybridization of raft culture of a cell line derived from a low-grade cervical neoplasm and containing episomal copies of HPV-31b. The dark spots indicate amplification of HPV viral genomes in highly differentiated koilocyte-like cells, similar to the differentiation-dependent amplification of HPV DNA that occurs* in vivo.

From Bedell, M.A., Hudson, J.B., Golub, T.R., Turyk, M.E., Hosken, M., Wilbanks, G.D., and Laimins, L.A. 1991. J Virol *65:2254–2260.*

Genetics and Biochemistry of Lipoprotein Lipase Deficiency

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.

ABNORMAL lipoprotein concentrations in plasma are commonly observed in the relatives of patients with early coronary disease, yielding various patterns of hypercholesterolemia and hypertriglyceridemia within families. Such complex phenotypes are thought to result either from the variable expression of a single-gene defect or the independent contribution of two or more genes. The genetic contribution to hyperlipidemia is further blurred by the fact that hormonal influences, diet, and habitus exert major influences on the regulation of lipid metabolism.

Our investigation of a large kindred yielded preliminary results in support of the multiple-gene hypothesis. In this pedigree one gene accounts for hypercholesterolemia, but hypertriglyceridemia depends on other factors. Among these, the gene encoding lipoprotein lipase (LPL) stood as a prime candidate.

LPL plays a key role in the metabolism of dietary and endogenous fat. Over 90 percent of dietary fat is hydrolyzed by this enzyme in an initial step controlling its delivery to peripheral tissues. The enzyme is secreted by mesodermal cells such as adipocytes and muscle cells, but acts at a distance from its site of synthesis. After secretion, it becomes anchored to the luminal surface of capillaries in extrahepatic tissues by an ionic interaction with heparan sulfate.

The enzyme needs apolipoprotein C-II as a cofactor to stimulate its catalytic activity. Also, dimerization is required. By binding to the surface of chylomicrons and very low density lipoproteins, LPL hydrolyzes triglycerides of intestinal or hepatic origin, releasing free fatty acids for cellular uptake where they can be used as fuel or reesterified for storage. Thus the enzyme plays a key role in the distribution of fatty acids among various tissues.

A host of factors regulate production of LPL, with stimulation by insulin, glucocorticoids, adenosine analogues, gastrin, and pancreozymin, and with inhibition by catecholamines, dibutyryl cAMP, and estrogens. While both the messenger RNA and the LPL gene have been characterized, little is known of the relationship between the enzyme's structure and its functional domains. These include a catalytic site and sites for lipid binding, heparin binding, and cofactor interaction. Other functional domains remain hypothetical.

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, recurrent acute pancreatitis, and eruptive xanthomas. Deficiency of the enzyme can be demonstrated in adipose tissue or in plasma after injected heparin induces its release.

The heterozygous state for LPL deficiency, by contrast, remains poorly characterized. Various reports document either normal lipids or moderate hypertriglyceridemia, and/or hypercholesterolemia, in relatives of deficient subjects. Could the heterozygous state, much more frequent than the homozygous, account for some form of the moderate to severe hypertriglyceridemia often noted in various clinical contexts?

We were able to address this issue by investigating the relatives of a subject with classical LPL deficiency. After identification of the molecular defect present in the homozygous state in the proband, and after demonstration of its functional significance through *in vitro* mutagenesis and expression, we determined carrier status with respect to this mutation among 126 relatives of the patient. We contrasted clinical and biochemical parameters collected on these subjects.

Our analysis revealed that the heterozygous state for this mutation indeed accounted for a common form of hypertriglyceridemia, accompanied by a sharp reduction of high-density lipoprotein cholesterol and subnormal levels of low-density lipoprotein cholesterol. These effects, however, were manifest only in subjects over 40. Our current hypothesis is that the heterozygous state imparts a latent, partial deficiency in the clearance of triglyceride-rich lipoproteins, which becomes manifest when triggered by other factors genetic or environmental.

As a consequence of these findings, LPL has become a major focus of research in our laboratory. We have initiated a systematic genetic and biochemical investigation of LPL in classical deficiency states and in a variety of clinical contexts that feature hypertriglyceridemia, including coronary heart disease, essential hypertension, diabetes, acute pancreatitis, and pregnancy-induced hypertension of diabetes.

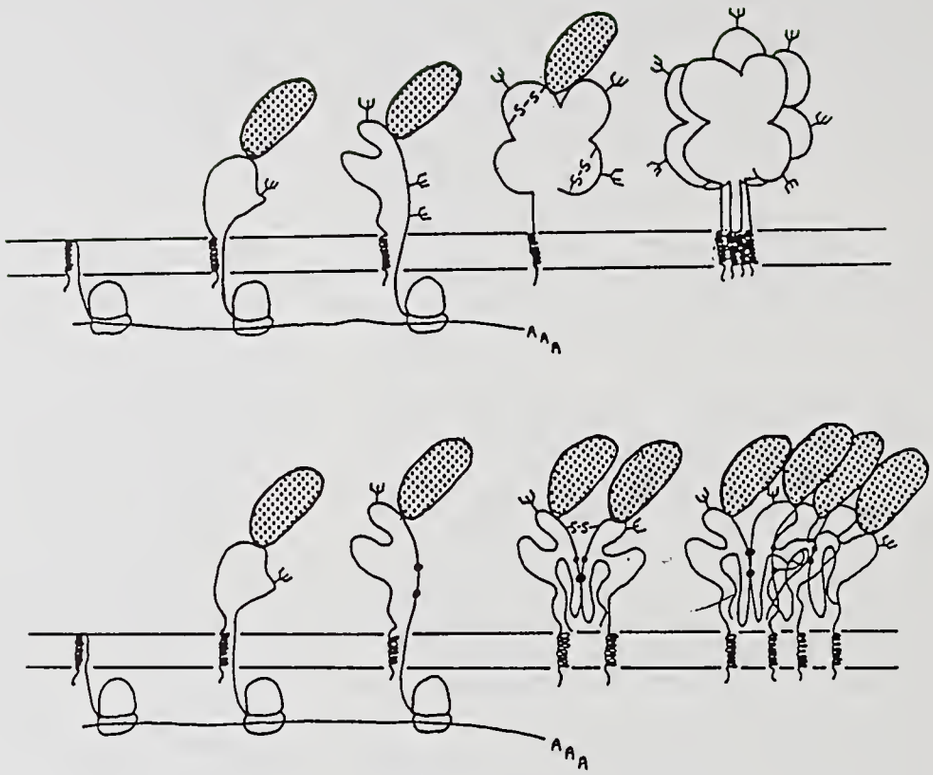
The identification of LPL mutations could clarify the pathophysiologic significance of hypertriglyceridemia in such conditions. Of yet greater interest to us, identification of naturally occurring mutations that impart a deficiency of functional enzyme will provide the basis for in-depth investigations of the functional domains involved and the structure-function relationships of LPL.

Our search has already led to the identification of multiple new mutations of the LPL gene. For each of those leading to amino acid substitutions, *in vitro* expression experiments have confirmed that the mutation produces immunoreactive but inactive enzyme. Most amino acid substitutions that we or others have so far identified cluster in a region that appears to harbor the catalytic domain of the enzyme.

This conclusion can be inferred from our recent report of the three-dimensional structure of a related lipolytic enzyme, pancreatic lipase. We predicted the identity of the three key amino acids defining the catalytic triad of LPL. Systematic amino acid substitutions were generated by *in vitro* mutagenesis at one site and expressed in cultured cells, in order to probe what aspect of the structure was altered by a naturally occurring variant. All substitutions led to the recovery of an inactive enzyme, suggesting that none could substitute for a glycine strategically located with respect to the catalytic site.

In further experiments in progress, we are probing other putative functional domains of the enzyme, particularly those regions that may be involved either with cofactor interaction or with heparan sulfate.

We have extended our investigation of molecular variation at the LPL locus by searching systematically for mutations of the gene among more than 130 unrelated individuals presenting a common form of hypertriglyceridemia in association with premature coronary artery disease, essential hypertension, hyperlipidemia, or pancreatitis in pregnancy. A variety of molecular variants have been identified on the basis of altered electrophoretic mobilities and are currently being characterized by DNA sequencing. We already know that a number of common silent mutations account for a significant proportion of the variants observed. However, conclusions with respect to the relative contribution of LPL defects to the pathogenesis of hypertriglyceridemia cannot be made at this stage of our ongoing study.



Two roles of the cellular chaperone protein GRP78-BiP in the endoplasmic reticulum (ER). This protein, a member of the stress (heat-shock) family, is a resident component of the ER and has peptide-binding activity. The top section illustrates the specific and transient association of GRP78-BiP (dotted oval) with a viral glycoprotein, hemagglutinin-neuraminidase, during its synthesis and folding in the ER. These entities provide an interesting model system, as GRP78-BiP is known to be released from the glycoprotein at a point in time prior to oligomerization. The bottom section illustrates the more stable association of GRP78-BiP with malfolded integral membrane proteins in the ER. If the protein contains a lesion that prevents native folding, then GRP78-BiP remains associated with it and normal oligomerization does not occur.

Research of Robert A. Lamb.

Structure and Replication of Influenza Virus and Paramyxoviruses

Robert A. Lamb, Ph.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. degree 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. Nine years ago he joined the faculty of Northwestern University.

ANIMAL viruses provide a unique tool with which to increase understanding of 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 viruses, influenza virus and the paramyxovirus SV5.

Influenza virus causes important diseases in humans and animals. It has tremendous socioeconomic consequences, because 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. Among these viruses are parainfluenza types 1–5, mumps, measles, canine distemper, Newcastle disease of chickens, and rinderpest of cattle, as well as SV5.

We are elucidating the wide range of mechanisms that these RNA viruses use to maximize the encoded proteins 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 (yielding a separate mRNA), and a coupled stop-start translation of tandem cistrons.

Influenza virus and paramyxoviruses (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 various properties of integral membrane proteins. Since these membrane proteins constitute the viruses' major antigenic determinants, knowledge about the structure of these proteins should enhance our understanding of their ability to act as immunological targets, thus aiding in developing rationally designed therapeutic agents and new means of vaccination.

We are investigating the mechanism by which integral membrane proteins are transported to the cell surface in the exocytotic pathway and are internalized from the surface by the endocytotic pathways. Our model systems include the seven integral membrane proteins encoded by influenza virus and SV5, three of which were discovered in our laboratory. These prototype membrane proteins, grouped as types I, II, and III, provide a diverse array of viral proteins that span the cell membrane once.

Intracellular Transport of Glycoproteins

We are determining how polypeptides are initially inserted into the endoplasmic reticulum and what signals are necessary for the proteins to interact with the lipid bilayer. To elucidate the rules for protein orientation in the bilayer is a prime objective. 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 cell 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 endoplasmic reticulum.

The cellular glucose-regulated protein GRP78-BiP is a member of the HSP70 stress family of gene products and a resident component of the endoplasmic reticulum, 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. The SV5 hemagglutinin-neuraminidase (HN) glycoprotein during its folding specifically and transiently associates with GRP78-BiP. The fact that this complex formation can only be detected before oligomerization of the immature HN molecules forms the native tetramer suggests that GRP78-BiP acts as a chaperone to promote correct folding of the molecule. Paramyxovirus infection of cells causes a transcriptional induction of GRP78-BiP mRNA, and our studies indicate that the flux of HN through the endoplasmic reticulum, which requires GRP78-BiP molecules for its maturation, causes a feedback that increases GRP78-BiP transcription.

To attempt to understand the nature of the specificity of GRP78-BiP with a protein, we constructed various altered HN molecules. The data indicate that HN contains more than one domain

capable of interacting with GRP78-BiP and that, without the binding of GRP78-BiP to the altered HN molecules, an increase in the accumulation of GRP78-BiP mRNA does not occur. Thus the transduction of the signal from the endoplasmic reticulum to the nucleus minimally requires the binding of GRP78-BiP to protein.

Internalization and Degradation of Glycoproteins

The SV5 HN glycoprotein is internalized extensively from the virus-infected cell surface and degraded in lysosomes. We are intensively examining the mechanism of 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 via 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 species targeting to lysosomes.

Virus Cation Channels

The influenza virus M_2 protein is a small (97-residue) type III integral membrane protein that forms a disulfide-linked tetramer. Circumstantial evidence based on the virus's sensitivity 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 has led to the suggestion that M_2 is an ion channel that alters the intracellular pH of the trans-Golgi network. Our direct experiments (in collaboration with Lawrence Pinto) of injecting M_2 mRNA into *Xenopus* oocytes and using a two-electrode voltage-clamp apparatus to measure surface currents indicate that, in the presence of M_2, an amantadine hydrochloride–sensitive cation channel conductance can be measured that has the characteristics of an H current, an electrical response previously observed in heart and neuronal cells.

We are beginning a structure-function analysis of this cation-specific channel to identify the residues involved in drug sensitivity and cation specificity. We are also investigating the possibility that the influenza virus NB glycoprotein, the paramyxovirus SH protein, and the human immunodeficiency virus type I *vpu* integral membrane proteins, all of which have M_2-like structures, have cation channel activity.

Virus Fusion Activity

We are also interested in the manner in which the F protein of paramyxoviruses causes cell fusion. Considerable evidence has implicated the hydrophobic amino terminus of the F_1 subunit (FRED) as being directly involved in membrane fusion, and we have shown that the FRED can interact stably with a lipid bilayer. Comparison of the amino acid sequences of many paramyxoviruses has revealed a high degree of amino acid identity in the FRED. To investigate the role of the conserved residues in mediating cell fusion, we have constructed many mutant F genes in which the invariant residue of FRED has been changed, in most cases for a similar residue.

Unexpectedly, in view of the remarkable amino acid conservation between paramyxoviruses, all of the mutant fusion proteins that are expressed at the cell surface are capable of causing cell fusion, which suggests that the residues serve some other function of F protein activity. Interestingly, when the highly conserved glycine residues in the FRED are changed to alanine residues—a change that can be predicted to make the FRED more α-helical in nature—the specific activity of the F protein in causing cell-cell fusion increases by several orders of magnitude. Thus we have engineered an even more potent fusogenic protein than the natural F.

The Enzymes Causing Thyroid Hormone Activation

P. Reed Larsen, M.D.—Investigator

Dr. Larsen is also Professor of Medicine at Harvard Medical School and both Director of the Thyroid Diagnostic Center and Senior Physician at Brigham and Women's Hospital. He obtained his bachelor's degree in English literature from Princeton University and his M.D. degree from Columbia University College of Physicians and Surgeons. He received postdoctoral training at Presbyterian Hospital, New York City, and the NIH and served for five years on the faculty at the University of Pittsburgh School of Medicine. He has been a member of the Harvard Medical School faculty for the past 16 years. Honors for his contributions to research on thyroid physiology and disease include the Van Meter-Armour Award and the Parke-Davis Distinguished Lectureship of the American Thyroid Association.

THYROID hormone is critically important in the development and regulation of metabolic processes. In humans the hormone is especially necessary for normal development of the central nervous system. Its importance is exemplified by the irreversible mental retardation that ensues if congenital hypothyroidism is not recognized and treated within the first few months of life. In infancy and childhood, thyroid hormone is absolutely required for normal growth.

Our laboratory is interested in learning how this pluripotent hormone produces its effects. The synthesis of thyroid hormone, thyroxine, is a tightly regulated process requiring oxidation of iodine and its incorporation into thyroglobulin, a thyroxine precursor. This protein is present in large quantities in the thyroid gland, and under regulation by pituitary thyroid-stimulating hormone, thyroxine is released into the bloodstream.

The thyroxine molecule has four iodine atoms distributed over its two-ring structure. In this form, the hormone is relatively inactive. To become fully potent, thyroid hormone must be deiodinated to form 3,5,3′-triiodothyronine. During this process, a specific atom of iodine is lost from the thyroxine molecule, and it is therefore termed T_3.

The process of thyroid hormone activation is thus one of great general interest. The enzymes that carry out this reaction are termed deiodinases, since they cause the removal of an iodine atom. There are two major deiodinases in the body. One is present in the liver and kidney and acts on thyroxine to form T_3, which then reenters the bloodstream. By this route T_3 is carried to other organs in the body, such as the skeletal muscle and heart, which do not have the capacity to produce it. The enzyme requires an intracellular cofactor for its function, which in turn requires adequate nutrition and the absence of stress for its production. In persons who have fasted or become ill, conversion of thyroxine to T_3 by this enzyme is markedly slowed, allowing thyroid hormone to fall to conserve energy.

A major focus of our laboratory has been to isolate and clone this deiodinase. The purification of this membrane-bound protein is made difficult by its insolubility in water and its inactivation by oxygen. To circumvent these difficulties, we have used a sensitive measurement for the enzyme's activity. We quantitate the release of radioactive iodine from an artificially synthesized analogue of thyroxine. This analogue is an even better substrate for the enzyme's deiodination process than is thyroxine itself.

Having a sensitive assay has allowed us to devise a strategy for cloning the enzyme without purifying it. For this purpose we have adopted the technique called expression cloning. In this process, one tries to identify sequences of DNA coding for a protein by measuring the capacity of the DNA to program a cell to make the protein. The strategy is labor intensive, but it assures an accurate sequence if successful. One common approach to expression cloning is to employ the egg (oocyte) of the African clawed toad (*Xenopus*) as a highly efficient factory for making proteins from the messenger RNA (mRNA) encoded by the DNA of interest. If this system is to be successful, an active form of the protein must be produced in the egg after it is injected with the mRNA from the tissue that expresses the enzyme. In the case of the deiodinase, injection of mRNA into *Xenopus* oocytes led to the appearance of the deiodinase in the egg proteins, indicating the feasibility of the approach.

To obtain the optimal starting material for preparing artificial DNA sequences to code for the deiodinase, we analyzed the effect of altering an animal's thyroid hormone status on the activity of this enzyme. If an animal is made hyperthyroid—i.e., given large quantities of thyroid hormone—the activity of the enzyme is increased and the amount of mRNA coding for the enzyme in liver and kidney is increased. Conversely, when an animal's capacity to synthesize thyroid hormone is blocked, the mRNA levels for the enzyme fall. Thus we have chosen hyperthyroid animals to use

as the starting source for preparing a DNA library for this purpose.

The preparation of cDNA from the liver of the hyperthyroid rat was size selected because studies showed that it was approximately 2,000 nucleotides in length. Many cDNA clones of this size were separated into pools of approximately 2,000 each. Then the normal process that is carried out in the cell was reproduced in the test tube. Messenger RNA was synthesized from each of the pools of these cDNA clones, injected into *Xenopus* oocytes, and the oocytes were analyzed three days later for the deiodinase activity.

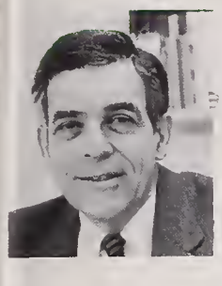
One pool produced a significant quantity of deiodinase and was therefore subdivided further. Pools of approximately 100 cDNAs were again studied by transcribing mRNA, and this was injected in turn into the *Xenopus*. By this technique a positive pool of about 100 of these clones was isolated. These were diluted and grown up as single colonies in a matrix arrangement. The RNA was prepared from pools of these clones from rows and columns, and an intersection was identified that corresponded to the clone that coded for the deiodinase. As anticipated, the clone was approximately 2,100 nucleotides in base pairs in length. Surprisingly, there was a group of three nucleotides, TGA (referred to as a codon), in the clone that corresponded to the signal for the protein synthesis machinery to stop. This codon, however, did not stop the synthesis of protein for this particular mRNA. We were able to show that in the case of this protein the stop codon TGA (UGA in the mRNA) coded for an extremely rare amino acid, selenocysteine. This amino acid is similar to the more common amino acid cysteine, except that selenium replaces the sulfur atom. Selenium is in the same category in the periodic table of chemical elements as is sulfur, but its chemistry differs from that of sulfur in a number of important respects. Only one other enzyme previously identified in animals and plants contains a selenocysteine. The role of this enzyme—glutathione peroxidase—is to reduce the peroxide that accumulates in various cells and damages cell membranes.

In the deiodinase, the selenium atom donates the electron that permits the removal of iodine from the thyroxine molecule to activate it. The chemical nature of the reaction—reduction of substrate and oxidation of the enzyme—is well known in cellular biochemistry. We were able to prove that this amino acid containing selenium is essential for the normal deiodination process, which explains the recent observation that a deficiency of selenium (which is present as a trace element in the earth's crust) causes hypothyroidism in experimental animals. We conclude that selenium is required for normal thyroid hormone action.

Our current research is directed at understanding why this particular mRNA is able to permit the suppression of the normal stop codon and therefore produce this unique enzyme. We are also investigating the biochemistry of the normal and mutated deiodinase molecule, so that we may understand more about its physiological function and deiodination mechanism.

Our laboratory continues to investigate the function of the nuclear protein to which T_3 binds and the DNA sequences that recognize it. With these two efforts, we hope to have a complete picture of how thyroid hormone produces its effects, beginning with the prohormone thyroxine and ending with the actual molecular mechanism by which thyroid hormone–responsive genes are regulated.

Cancer and Genetic Modification of Biologic Control



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 the 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 within an organism is far too delicate and important a process to be left to chance. Rather, as with all biologic processes, it is subject to a stringent set of rules that are programmed into the genetic makeup of the organism. Genes form the basis for controlling growth within an organism, setting the parameters that allow the liver to take the shape it does or the kidney to assume its particular size and function. Genes establish the rules that ensure that an organ grows in an orderly fashion and reaches a prescribed and limited size. Growth can thus 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, a 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 at the wrong place or, indeed, even cause it to make the wrong product. The set of genes whose damage (or mutation) can give rise to cancer are often just those genes 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 and, in particular, the control processes that operate in the living organism. Our work has been considerably advanced by the technique of introducing active oncogenes into the hereditary makeup of special strains of laboratory mice. These "transgenic" mice carry oncogenes created in the laboratory, pass these cancer-causing genes on 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. These experiments have taught us that some cancers can be caused by certain 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, a process that requires several activating events.

Host Defenses Against Cancer

Although transgenic mice are very useful in analyzing the action of oncogenes, they are also useful in exploring the host defense mechanism that can be 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 example, 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 at best poorly understood, but important recent discoveries in the field of 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. They can act as growth factors signaling target cells to divide, or they can trigger genetic programs that cause a target cell to change, mature, or even assume a new function.

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 given the name

IL, stemming from a time when they were called "interleukins," or agents that mediated signals between white blood cells, or leukocytes.) IL-4 has several cell-signaling functions that have been recognized by studies in tissue culture. 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 have a major role in the orderly development of antibody-producing cells, the so-called B cells.

To learn what role, if any, these interleukins play in modulating the body's response to cancer, we devised a powerful test for their function as antitumor agents. We planted the gene for IL-4 in a tumor cell that allowed its high-level production. The IL-4-producing tumor grew well in a tissue culture dish, suggesting that IL-4 did not directly block growth of the tumor cell. However, when the IL-4-producing tumor was transplanted to a host mouse, the tumor growth was completely blocked. Evidently IL-4 was eliciting a host defense mechanism that prevented tumor growth. We have extended this observation to many different types of tumor cells under many conditions and have obtained similarly dramatic results.

Further studies of the antitumor action of IL-4 have allowed us to identify two particular cell types that may be mediating this antitumor effect. One of these is the eosinophil, a cell that is involved in many allergic responses. The second candidate mediator cell is the macrophage, a scavenger cell concerned with many cell-killing functions. Our most recent work has focused on the particularly esoteric eosinophil, a cell thought to be involved in defense against parasitic diseases and in certain allergic disorders. Studies using specific reagents now convince us that this peculiar cell is the major mediator of the antitumor effect.

We would also like to know whether a small increase in the production of IL-4 can influence an organism's ability to develop or defend itself against malignancy. Accordingly, we made transgenic mice that carry extra copies of the IL-4 gene to ensure that their lymphocytes produce large amounts of this agent. Creating these mice proved quite difficult, since high concentrations of IL-4 have an extremely disruptive influence on the murine immune system. We solved this problem by creating a graded set of transgenic mice that yielded reduced levels of IL-4. By means of special regulatory gene signals, we were able to create mice that produced IL-4 in minimally disruptive amounts. From these animals we have learned that IL-4 has a profound influence on the immune cells that develop in the thymus gland, a key organ in the immune system. But we have also seen that this lymphokine can induce a serious allergic inflammatory response, suggesting that we will have to deal with this side effect before considering the therapeutic uses of this agent.

We have expanded these studies to view the actions of other lymphokines that, theoretically, might influence tumor occurrence. Our preliminary studies indicate that IL-7 may provoke a tumor-inducing response, clearly indicating that different growth signals can have dramatically different effects in the organism.

Finally, although this brief account cannot cover the entire range of our studies, we have turned to a third element (the so-called *kit* ligand), which has a profound influence on the development of cells during the formation of the embryo. This factor influences the development of cells that contribute to the formation of blood, of pigment-producing cells, and of eggs and sperm. In other words, if a mouse cannot make this element, it is sterile, albino, and has a profound anemia. Because we feel that the developmental pathway a cell follows strongly influences its propensity to undergo a malignant change, we are studying this factor to learn how it influences this important process in the mouse.

Molecular Genetic Investigation and Therapy for Inborn Errors of Metabolism

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.

OUR laboratory has focused on molecular genetic investigations of the enzyme methylmalonyl CoA mutase (MCM) and its deficiency state in humans, methylmalonic aciduria. This enzyme is required for the metabolism of propionate, which can be produced by catabolism of certain amino acids or absorbed from fermentation products in the intestines.

We have cloned genes for human and mouse MCM, identified mutations that give rise to notable phenotypes of methylmalonic aciduria, and demonstrated correction of the metabolic defect in cultured cells from methylmalonic aciduria patients by gene transfer of a normal MCM gene. This last experiment is prescient of somatic gene therapy in which patients with inborn errors of metabolism might be treated by introducing a normal gene into somatic cells to perform the function of the inherited mutant genes.

Successful somatic gene therapy involves more than introducing a normal gene into genetically defective cells. It will be necessary to achieve proper compartmentalization and regulation of the recombinant enzyme, substrates, cofactors, and reaction products; to determine which organs and somatic cells represent the best possible targets for gene transfer, how many cells need to be transformed, and how the metabolic capacity and phenotypic effect of reconstituted cells may be maximized; and to address a variety of clinical issues to ensure that clinical trials embody the greatest potential benefit and entail the least possible risk. It is also necessary to adhere to the regulatory and review process that has been established to monitor human gene therapy.

Our studies are directed at attaining an understanding of the structure and function of MCM; of its role in homeostasis, nutrition, and pathology; and of the multifaceted issues required to propose gene therapy for methylmalonic aciduria. In the past year this has involved basic research concerning MCM and gene transfer as well as the establishment of a clinical foundation for gene therapy trials.

Role of Methylmalonyl CoA Mutase in Health and Disease

The clinical expression of MCM deficiency presents several enigmas. Why is this disorder lethal? Why are the symptoms intermittent? Why are some individuals with persistent methylmalonic aciduria clinically normal? We have described a variety of mutations in the gene for MCM. Some give rise to completely inactive enzymes, others to enzymes that show limited activity in the presence of vastly excessive cofactor, and an interesting enzyme that is nonfunctional in most cells but active in heterologous combination with some other mutant enzymes. We will study whether the residual activity expressed by mutant genes correlates with the different phenotypic forms of methylmalonic aciduria or whether other factors are involved in this pleomorphism.

We are studying propionate metabolism in various normal and genetically defective cells to investigate the role of MCM in homeostasis. Our studies suggest that under normal conditions the rate of propionate metabolism is not limited by the amount of the enzyme. Barely detectable levels of MCM can sustain maximal levels of propionate metabolism, and the addition of more enzyme to normal cells has no effect on this metabolic activity.

Interestingly, the level of propionate metabolism in cultures containing few cells with reconstituted enzyme is similar to that of a population of completely normal cells, indicating exchange of intermediate metabolites between cells. The exchange appears to require cell contact; it does not occur between cells separated by a fluid phase. This suggests that gene therapy targeting a small population of cells may have a significant phenotypic impact, though therapy may have to be targeted to organs damaged by the accumulation of organic acids, particularly the liver, rather than more accessible sites such as bone marrow.

We have also observed that the ability of a colonic epithelial cell line (CACO-2) to metabolize propionate increases dramatically as the cell differentiates in culture. This suggests that propionate metabolism may be a differentiated function of intestinal epithelium. There are several reasons to think that the intestine may play a critical role in propionate homeo-

stasis and that the fermentation and absorption of propionate may be an important factor regulating organic acid homeostasis.

Basic Methods for Gene Transfer

We have focused on the liver and the intestines as targets for somatic gene therapy. In an ongoing collaboration with Savio Woo (HHMI, Baylor College of Medicine), we have been involved in developing a scheme for hepatic gene therapy that involves harvesting and cultivating hepatocytes, transducing these cells with recombinant retroviral vectors, and returning them to the host via hepatocellular transplantation.

In collaboration with Susan Henning (Department of Pediatrics, Baylor College of Medicine), we have developed methods for transducing genes into intestinal epithelium. This involves instilling recombinant retroviruses into isolated intestinal segments in the presence of agents to increase infectivity, reduce mucus, and increase the number of dividing crypt cells. These experiments demonstrate that it is possible to transduce crypt cells, though considerable work remains to achieve more efficient transduction.

Clinical Foundation of Somatic Gene Therapy

We have also begun to establish a clinical foundation for somatic gene therapy by developing a clinical trial of hepatocellular transplantation for acute hepatic failure in children. We will use hepatic cells transduced with recombinant retrovirus expressing a marker gene (in collaboration with Savio Woo and others at Baylor College of Medicine). Hepatocellular transplantation, an essential of proposed schemes for hepatic gene therapy, has never been attempted in humans. We are establishing surgical and clinical methods for such transplantation. Use of marker genes will facilitate assessment of cellular engraftment and of techniques for transducing recombinant genes into hepatic cells.

The development of this protocol has focused our attention on the myriad clinical issues involved in human trials, including proper patient selection, obtaining meaningful informed consent, ensuring adequate follow-up, adopting methods that are approved for clinical use and meet the standards of quality control required for clinical practice, satisfying guidelines for containment, and involving various health professionals and the public in this exciting therapeutic frontier. Many of these issues have become research projects in themselves as we have tried to establish a formal basis for clinical trials of gene therapy.

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, the 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 several distinct subtypes of adrenergic receptors: α_1, α_2, β_1, and β_2. 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. They 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 them 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 recently developed 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 recently isolated the genes for all of the known adrenergic receptors (α_1, α_2, β_1, β_2), 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 physiologic 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 four main 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 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 in 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. A wide variety of other circumstances that regulate receptor properties were also identified, including aging, congestive heart failure, β-blocker therapy, thyroid hormones, steroid hormones, and neoplasia.

Our recent research is increasing our understanding, in molecular terms, of how the receptors become functionally desensitized. We have recently discovered a new enzyme, β-adrenergic receptor kinase (βARK), which modifies the structure of the receptor by introducing a phosphate group. This modification may provide the basis for receptor desensitization. We are currently studying how such phosphorylation of the receptors by βARK and related enzymes modifies their function.

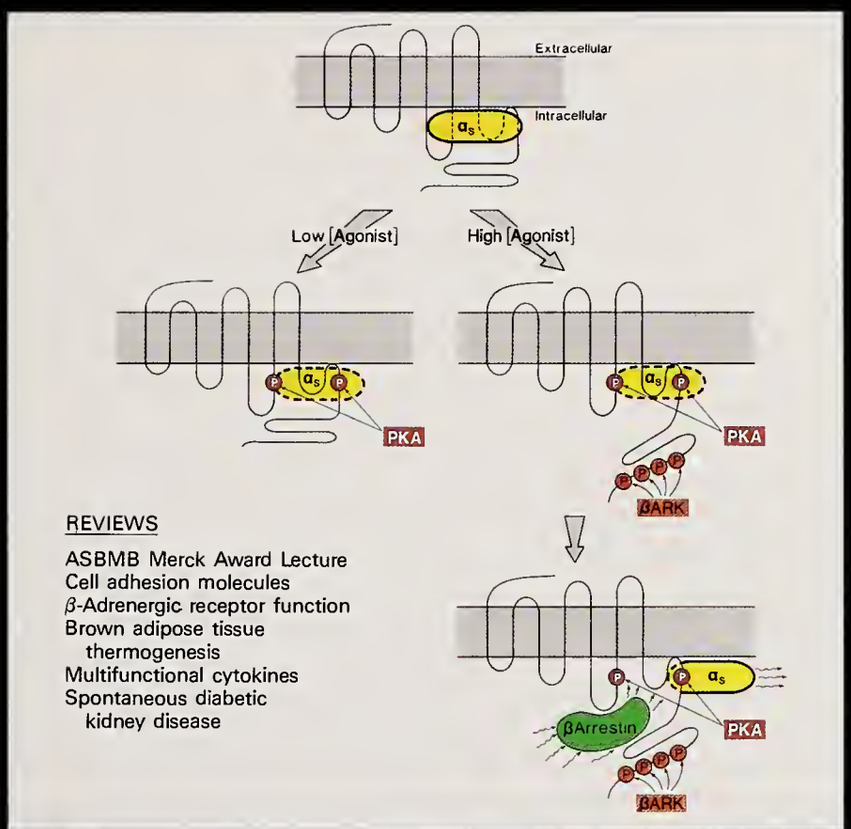
The implications of such fundamental research on receptors for clinical medicine are profound. Elucidation of the detailed structure of the receptors will allow the precise design of more potent and specific drugs. Unraveling of the molecular basis of desensitization will allow the development of strategies 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.

Opposite: Desensitization of the β-adrenergic receptor. Many pharmacological agents lose their effectiveness upon repeated or prolonged administration. The molecular mechanisms responsible for this "desensitization" may be analogous to those that underlie the loss of responsiveness observed in cell culture systems to hormones that activate β-adrenergic receptor molecules. Activation of these receptors promotes their rapid association with the α-subunit of the stimulatory G protein (α_s). Depending on the concentration of the activating hormone, other processes are also triggered that eventually disrupt this association, including phosphorylation of the receptor molecule by the cAMP-dependent protein kinase (PKA) and the β-adrenergic receptor kinase (βARK). Disruption of receptor-α_s coupling is further ensured or maintained by interaction of the phosphorylated receptor with the recently identified cytosolic protein β-arrestin.

From Hausdorff, W.P., Caron, M.G., and Lefkowitz, R.J. 1990. FASEB J *4:2881–2889.*

THE FASEB JOURNAL

Official Publication of the Federation of American Societies for Experimental Biology



REVIEWS

ASBMB Merck Award Lecture
Cell adhesion molecules
β-Adrenergic receptor function
Brown adipose tissue thermogenesis
Multifunctional cytokines
Spontaneous diabetic kidney disease

Volume 4, Number 11 August 1990

Genetic Control of Pattern 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, Assistant Professor of Biology at the Massachusetts Institute of Technology, and Assistant Molecular Biologist at the Massachusetts General Hospital, Boston. She received her M.Sc. degree from the University of Freiburg, where she worked with J. 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 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.*

OUR laboratory is interested in the genetic mechanisms that control the establishment of polarity during embryogenesis. In *Drosophila*, basic information about the "coordinates" of the embryo is supplied to the egg cell during its maturation in the mother. Mutations in maternal genes have led to the identification of a small number of genes that are required for the establishment of dorsal-ventral and anterior-posterior polarity in the embryo. In these mutants the lack of a particular maternal gene product leads to a lethal phenotype in the embryo.

Three independent morphogenetic systems are required for the establishment of pattern along the anterior-posterior axis. The anterior system controls the development of head and thorax, the posterior system determines the abdominal region, and the terminal system is required for the most anterior and posterior structures. The anterior and posterior systems act through factors that are localized to the respective poles. For the anterior group of genes, the localized factor is the mRNA product of the gene *bicoid*. The other anterior genes are involved in the localization of *bicoid* RNA. Localization of the *bicoid* RNA is necessary for the formation of a concentration gradient of bicoid protein. This gradient along the anterior-posterior axis determines thresholds for the expression of zygotic target genes. Thus high levels of bicoid promote head formation, while lower levels promote thorax formation.

We have concentrated on dissecting the pathway that leads to normal development of the posterior region of the embryo. Nine maternal genes have been identified that set the basic posterior pattern. These genes, referred to as the posterior group, share the abdominal phenotype: homozygous mutant females produce offspring that lack abdominal segmentation. Only the *nanos* gene, however, encodes the localized signal required for the development of the abdomen, while seven genes—*oskar*, *vasa*, *tudor*, *valois*, *staufen*, *cappuccino*, and *spire*—are required for the localization of *nanos*.

These genes, in addition to the abdominal phenotype, show an effect on germ cell formation: mutant embryos do not form a specialized posterior pole plasm that normally contains the polar granules, and hence these embryos lack pole cells, the precursors of the germline cells in *Drosophila*. One gene, *pumilio*, is required for the distribution of the nanos protein. Our molecular analysis of the posterior group is aimed at determining how *nanos* and various 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 cytoplasmic zone at the posterior pole is the source of an abdominal signal and that embryos derived from females mutant for any one of the posterior group genes lack a common signal in the abdominal region. Depending on the gene affected, mutant embryos are deficient in the synthesis, storage, or transport of the signal.

The gene *nanos* is central to the genetic pathway leading to abdomen formation. We have isolated and cloned the gene and have shown that it encodes the posterior signal. During oogenesis, *nanos* mRNA is transcribed and becomes localized to the posterior pole of the oocyte. The central role of *nanos* for posterior pattern formation is demonstrated by the finding that *nanos* transcript synthesized *in vitro* from a *nanos* cDNA rescues the abdominal segmentation phenotype not only of *nanos* mutant embryos but also of all posterior group mutants so far tested. Since seven of the nine mutants (except *nanos* and *pumilio*) also lack pole plasm and fail to localize *nanos* RNA, we can conclude that in pole plasm–defective mutants the abdominal segmentation defect is a consequence of a lack of localized *nanos* activity. In future experiments we will determine cis-acting sequences within the *nanos* mRNA,

and the trans-acting factors required for *nanos* RNA localization.

We are just beginning to understand the process through which the nanos protein affects the development of abdominal structures. The gene appears to inhibit the activity of another maternally derived protein, encoded by the *hunchback* gene. Maternally derived hunchback protein seems to act as a repressor molecule; it blocks the transcription of genes within the embryo. The *nanos* gene acts negatively on the mRNA stability and/or translation of the maternally provided product of *hunchback*, such that the maternal *hunchback* transcript and protein product are absent from the posterior of the embryo. Thus, by interfering with *hunchback* activity, *nanos* indirectly promotes activation of the first tier of embryonic genes required for abdomen formation.

Distribution of the Posterior Signal

The *pumilio* gene is required for the distribution of the abdominal signal, encoded by the *nanos* gene, during early embryogenesis. It therefore provides an excellent opportunity to study the redistribution of localized cytoplasmic determinants. Initial cytoplasmic transplantation experiments indicated that in the absence of the maternally provided *pumilio* gene product, the abdominal signal *nanos* is properly localized to the posterior pole but does not reach its target, the prospective abdominal region. *In situ* hybridization studies have demonstrated that *nanos* mRNA is localized to the posterior pole plasm in wild-type embryos and that its distribution is unaffected by a *pumilio* mutation. This suggests that the *pumilio* gene affects the distribution of the nanos protein.

We have cloned the *pumilio* gene and have shown that it encodes several transcripts that are present throughout development. In early embryos, *pumilio* transcripts are localized to the posterior pole. This observation is consistent with the possible interaction between *pumilio* and *nanos* gene products. The localization of the transcript is disrupted by the same set of mutants that disrupt *nanos* localization, and it is possible that the same mechanism of RNA localization acts on both RNAs.

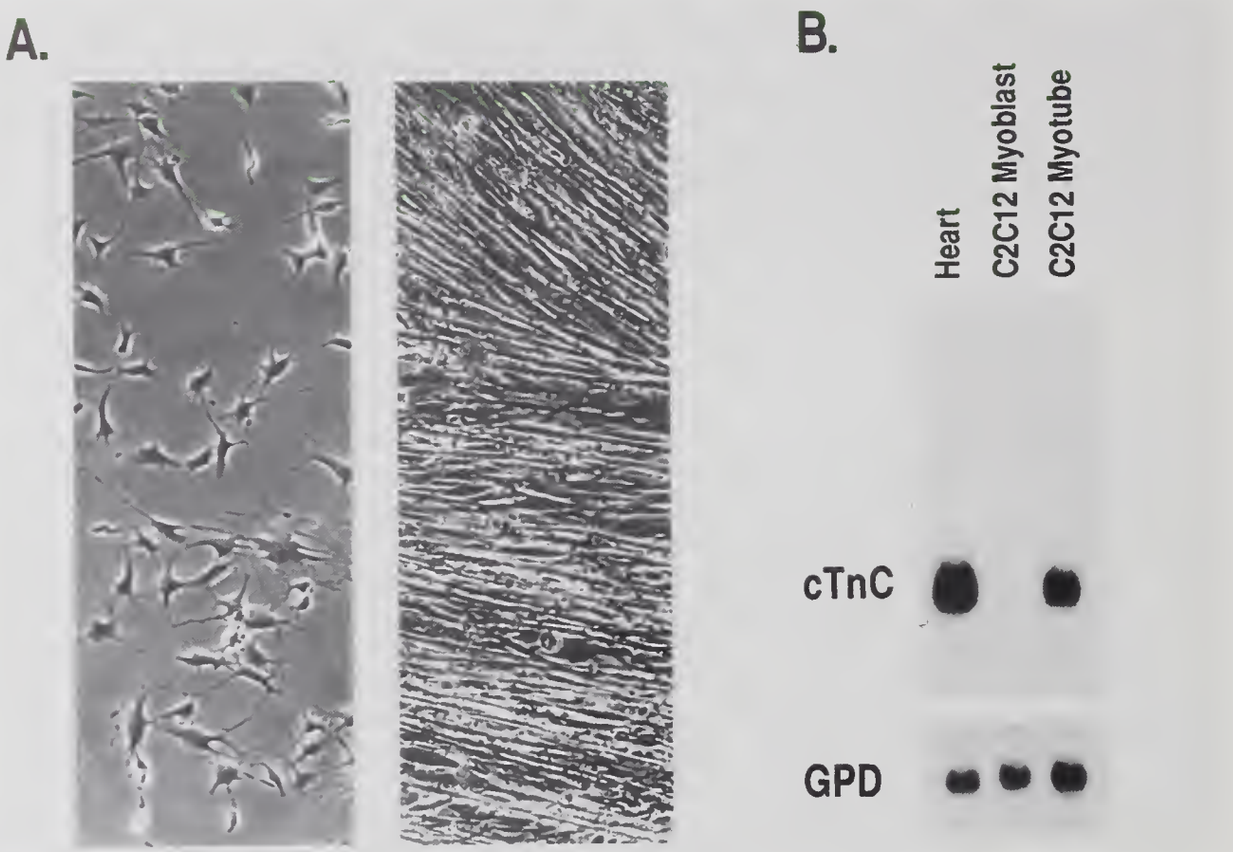
Assembly of the Pole Plasm

The remaining seven genes in the posterior group are associated with pole plasm defects. We are interested in their role in pattern formation and thus in *nanos* RNA localization. We are also very excited about the potential for learning more about the origin of the germline.

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 the zone actually contains specialized cytoplasm packed with mitochondria and numerous donut-shaped organelles that do not occur anywhere else in the embryo. These organelles, called polar granules, must be present for the formation of pole cells, the precursors of the germ cells. 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 pole cells has provided a new avenue for research.

Our hypothesis is that the seven remaining posterior group genes do not have a direct role in transmission of the abdominal signal. Instead, they provide the structural framework that localizes *nanos* mRNA in the pole plasm. This framework may be the polar granule or components thereof. 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 are optimistic that new information about the assembly and composition of the pole plasm and polar granules will lead to greater understanding of their various functions in the early embryo. 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.

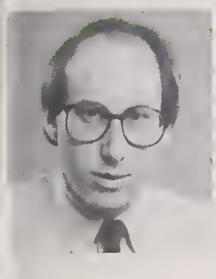


Developmental regulation of cTnC gene expression in vitro. *A: Differentiation of C2C12 myoblasts into myotubes induced by serum deprivation. Cultures of C2C12 myoblasts are shown on the left. After 48 h of serum deprivation, more than 95 percent of the myoblasts have differentiated into myotubes (right panel). Magnification, × 54.*

B: Northern blot analysis of RNA samples isolated from murine BALB/c cardiac tissue (marked Heart*), undifferentiated C2C12 myoblasts* (C2C12 Myoblast), *and differentiated C2C12 myotubes* (C2C12 Myotube) *hybridized to the murine cTnC cDNA probe. The bottom panel shows the same filters hybridized to a glyceraldehyde-3-phosphate dehydrogenase (GPD) probe.*

From Parmacek, M.S., and Leiden, J.M. 1989. J Biol Chem *264:13217–13225.*

Regulation of Gene Expression During Cell 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 expression patterns of these differentiation- and activation-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 understand better the molecular mechanisms that regulate gene expression during normal and pathologic development.

Regulation of Human T Cell Receptor (TCR) Gene Expression During T Cell Development

Human T cells recognize foreign antigens, such as virus-infected cells and tumor cells, via specific cell-surface TCR molecules. T cells can be divided into two subsets, based on their expression of two distinct types of antigen receptor molecules. The majority of circulating, peripheral blood T cells (including all helper and cytotoxic T cells) express the TCR α/β heterodimer; a small but distinct T cell subset of unknown function expresses the γ/δ TCR. These α/β and γ/δ T cells appear to develop as separate lineages during thymic ontogeny. During the past several years, my laboratory has been interested in the molecular mechanisms that regulate the expression of these TCR genes during T lymphocyte 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 TCR α and β genes in T cells and to function equally well in both mouse and human cells. Despite these similarities, they display a number of distinct molecular properties. For example, the TCR α enhancer functions only in α/β T cells and is equally active on a number of different promoters; the TCR β enhancer is quite active in both α/β and γ/δ T cells, displays low-level activity in B cells, but is inactive in nonlymphoid cells. The identification and localization of the human TCR enhancers led us to propose that certain T cell tumors previously 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 proto-oncogenes.

Our more recent studies have focused on identifying and characterizing the enhancer DNA sequences and the nuclear proteins bound by these sequences 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 are 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 nuclear protein-binding site that was identical to the previously described cAMP response element (CRE).

To clarify the function of the CRE in regulating T cell gene expression, we have cloned two novel CRE-binding proteins, CREB-2 and CREB-3, which specifically bind to the TCR α CRE and to CREs from several other eukaryotic promoters. Both of these 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.

More recently 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. Finally, in collaboration with Stuart Orkin (HHMI, the Children's Hospital, Boston) we have identified a new 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 normal T cell development and activation, as well as in T cell tumors.

Molecular Characterization of the Cardiac and Skeletal Muscle Troponin C Genes

Normal heart muscle contraction depends on the phenomenon of excitation-contraction coupling—i.e., the transmission of the electrical depolarization signal into the biochemical events involved in muscle cell contraction. Cardiac troponin C (cTnC) is the calcium-binding subunit of the muscle fiber that regulates excitation-contraction coupling. Cardiac or skeletal muscle contraction is initiated by electrical depolarization of the muscle cell membrane, resulting in a dramatic rise in intracellular calcium, which in turn binds to the troponin C molecule, causing a conformational change in the structure of the polypeptide. This signal is transduced across the muscle fiber, facilitating the formation of actin-myosin cross bridges, and ultimately resulting in the generation of tension in the muscle fiber.

Cardiac and skeletal muscle each express a different form of troponin C. These two forms display markedly different biophysical properties that, in part, account for the different properties of cardiac and skeletal muscle contraction. To clarify the function and regulation of the cTnC and skeletal muscle troponin C (sTnC) molecules, we have recently cloned cDNAs and genes that encode both of these proteins. Structural studies have revealed that these genes belong to a common multigene family that contains at least 8–10 closely related members. Although the two proteins are highly related, we have identified three areas that display significant structural diversity and that may account for the functional differences. Studies of the expression of the two genes have shown that the cTnC gene is expressed only in embryonic (not adult) fast skeletal muscle and in adult cardiac and slow skeletal muscle, while the fast sTnC gene is expressed only in embryonic and adult fast skeletal muscle. Although neither gene is expressed in cultured myoblasts (muscle cell progenitors), the expression of both is markedly induced following chemical differentiation of these myoblasts into embryonic myotubes.

Recently we have focused on identifying the genetic elements that regulate the tissue-specific expression of the two TnC genes during muscle cell development. These studies have demonstrated that the cTnC gene contains two distinct tissue-specific transcriptional enhancer elements: 1) a first-intron skeletal muscle–specific enhancer that is required for expression of the cTnC gene in embryonic skeletal myotubes and 2) an upstream cardiac-specific enhancer that is required for high-level cTnC gene expression in cardiac myocytes. Interestingly, the sTnC gene also contains a first-intron enhancer that is active in skeletal but not cardiac muscle cells. Ongoing studies are designed to dissect the molecular requirements for cTnC and sTnC enhancer function and to clone and characterize the enhancer-binding proteins that control the expression of these three enhancers.

More recently we have developed a method for introducing foreign genes directly into heart muscle cells in living animals. These inserted genes are expressed in the heart muscle for as long as six months. This simple technique, which involves injection of recombinant genes directly into the beating hearts of rats and mice, holds promise for the treatment of a number of human acquired and inherited cardiovascular diseases.

Preliminary Events in Olfaction and a Tissue-Specific snRNP Protein

Michael R. Lerner, M.D., Ph.D.—Associate Investigator

Dr. Lerner is also Associate Professor in the Department of Internal Medicine at Yale University School of Medicine. He obtained his B.A. degree in chemistry from the University of Pennsylvania. He received his M.D. degree from Yale University as well as the Ph.D. degree for work done with Joan Steitz on small nuclear ribonucleoproteins (snRNPs). He had an internship 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 Wilson S. Stone Award and the Lee C. Howley, Sr., Prize for research in arthritis.

IN human beings, olfaction is primarily a hedonistic sense; among other things, it provides most of the enjoyment derived from eating. For other animals, however, olfaction is important for finding mates and food and avoiding enemies and noxious compounds. To study the mechanisms behind olfaction, we are using as a model system the moth *Manduca sexta*.

Male moths locate females by their keen ability to detect sex pheromones released by females up to several miles distant. The "nose" of a moth is its antennae. The antennae are covered with fine hair-like processes called sensilla. The surface of a sensillum is composed of chitin perforated by numerous small pores, which open into tubules leading into a receptor lymph. Inside each sensillum and surrounded by the receptor lymph are the cilia from olfactory receptor neurons.

For an air-breathing animal to detect an odor, three events must take place. First, the volatile chemical in question, which is often quite hydrophobic, must cross an aqueous-based medium—in the case of moths, the receptor lymph—to reach the cilia. A perfect example of such an odorant is bombykal, a 16-chain hydrocarbon with an aldehyde group at one end, which is a sex pheromone for *Manduca*. Second, the chemical must interact with its receptor. Third, the chemical, or at least its action, must be terminated. This last point is particularly important in the case of pheromones, where real-life working concentrations are subpicomolar, necessitating that noise be kept to a minimum.

The receptor lymph is inundated with a 15-kDa protein called pheromone-binding protein (PBP). It is widely hypothesized that PBP is involved in the translocation of pheromone from the pore tubules to the membranes of the cilia. We have been characterizing this binding protein in *Manduca* and are currently working on its developmental regulation. To facilitate these studies we have cloned and sequenced the cDNA encoding it.

If PBP is responsible for solubilizing and transporting pheromone to pheromone receptors, how do general odorants get to their receptors? We have found that the olfactory sensilla for detecting general odors have two proteins that are related to PBP. These proteins, which are expressed in both males and females, may be general odorant carriers. Recently we cloned and sequenced cDNA coding for both of these proteins, which have been named GOBPs (general odorant-binding proteins).

Odor inactivation could be the result of any of several processes, ranging from destruction of the chemical in question, once it has interacted with a receptor, to desensitization of the receptor. A developmentally regulated antenna-specific aldehyde oxidase (AOX) has been identified. This 150-kDa enzyme has a strong preference for bombykal as a substrate as opposed to other aldehydes. As a result of the AOX, the half-life of a molecule of pheromone, once it enters a sensillum, is less than 1 millisecond.

A Tissue-Specific snRNP Protein

Messenger RNA splicing is a fundamental process in eukaryotic cells. The precise removal of intervening sequences (introns) and the rejoining of exons to form an mRNA encoding a specific protein are in some ways analogous to making a movie. First many scenes are taken. The primary RNA transcript made from a DNA template is akin to this unprocessed film. Like the film, RNA is processed in several ways. For example, a modified nucleotide cap structure is placed at the beginning of the RNA transcript—just as a title and headings are placed at the beginning of the film. Also, a strip of nucleotides, the poly(A) tail, is added to the 3′ end of the RNA—as credits are added at the end of the film. One of the most interesting and basic things that happens to an RNA transcript during its maturation is the excision of introns and the connection of exons. Again, as in movie making, the film that has taken many hours to shoot must be edited so that only the most important two hours remain. The cellular components that perform these splicing reac-

tions are called snRNPs (small nuclear ribonucleoproteins; see the report in this volume by Joan Steitz, HHMI, Yale University). As might be expected for such a fundamental process as splicing, the individual components of the snRNPs, their snRNAs and proteins, are generally present in all tissues.

Usually when snRNPs splice a given pre-mRNA, there is only one possible final product. However, this is not always the case, particularly with regard to RNA processing in the central nervous system (CNS). Imagine a movie with more than one potential ending, or even alternative scenes in the middle, so that individual viewers could make their own choices. Alternative mRNA splicing is a precise process by which different cells, starting with the same primary RNA transcript, make related, but distinct, mature mRNAs. At least one snRNP protein is apparently present in a tissue-specific manner, occurring mainly in brain and to a much lesser extent in other tissues such as heart, adrenal gland, olfactory epithelium, and pigment cells. This protein, called N, has been identified and cloned in collaboration with Susan Amara (HHMI, Yale University). Although N is expressed predominately in the CNS, it is only present in neurons and not in the more abundant nonneuronal cells, such as glia. Because of the restricted tissue and cell expression of N, the genomic sequence controlling its expression is of interest. Recently this sequence has been obtained and is now being investigated.

Structural Determinants of α-Globin Gene Expression

Stephen A. Liebhaber, M.D.—Associate Investigator

Dr. Liebhaber is also Associate Professor of Human 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, hematology, and molecular biology at four other universities. As a postdoctoral fellow with David Schlessinger at Washington University, Dr. Liebhaber examined ribosomal RNA processing, and with Y. W. 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.

EXPRESSION of a eukaryotic gene is a multistep process. Specific regions of chromatin must be activated and transcribed, transcripts processed, and mature mRNA exported to the cytoplasm for translation into protein. Gene expression can be controlled at any or all of these steps. 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 oxygen transport 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. Defects in either set result in an imbalance of expression and consequent anemia: α- or β-thalassemia. Thalassemias result from almost 100 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 synthesis (transcription) and an unusual stability of the mature globin mRNA.

Another interesting characteristic is that 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, unusual mRNA stability, and clearly defined pattern of developmental switching are areas of special focus in our laboratory.

The loss of α-globin gene 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 a series of patients with a rather unusual form of α-thalassemia in that their silenced α-globin genes have entirely normal structures. When these genes are isolated and reintroduced into cells, they function as well as normal ones. The cause of the abnormality in these patients appears to lie in a region of DNA outside the genes themselves.

Through extensive DNA mapping, we have detected large deletions outside the α-globin cluster in three α-thalassemic individuals. One of these ectopic deletions is separated from the α-globin gene by at least 30,000 bases. By comparing the maps of each deletion, a region of common overlap is noted. Since a similar set of signals critical for transcriptional activation has been identified adjacent to the β-globin gene cluster, one can speculate that such signals serve coordinately to activate the expression of both 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 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 introduced the human ζ- and α-globin genes into fertilized mouse eggs to generate transgenic mice. The red cells of these mice appropriately express the human transgenes during development. In the embryonic period, there is 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 human transgenes contain the necessary information for appropriate developmental control and 2) the factors responsible for developmental switching in the red cell have been sufficiently conserved during evolution to substitute in the control over the human transgenes. By generating transgenic mice that carry human ζ- and α-globin genes with specific alterations, and by studying their pattern of develop-

mental expression, it should be possible to arrive at a definition of the signal(s) critical to globin gene switching.

Selective stabilization of globin mRNA is essential to its accumulation in terminally differentiating red cells to levels of 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 codon substitution (CAA for UAA) at the termination of the normal 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 mRNA is unstable when the gene is expressed in an erythroid cell line, but normally stable when expressed in nonerythroid cell lines. This suggests that the stability of globin mRNA may depend upon interaction with one or more erythroid-specific factors.

By making specific alterations in the structure of the α-globin genes prior to expression, it has been possible to demonstrate that the region within the α-globin mRNA critical to its stability is located in a segment of the 3′-nontranslated region just past the translation termination codon. In future experiments the critical cellular factors or structures mediating this response, and the mechanism involved, will be further defined.

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 dissertation research was done with Matthew Meselson at Harvard University, where she began her work 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 employed by cells 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 highly conserved proteins, the heat-shock proteins (HSPs). These proteins are induced in response to temperature elevation and a wide variety of other stresses. This remarkable response is the most highly conserved genetic induction known, 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, with 40–50 percent amino acid identity between the proteins of human cells and bacterial cells.

Much indirect evidence suggests that the HSPs are instrumental in protecting cells and organisms from lethality at extreme temperatures. For example, pretreatments at moderately elevated temperatures, which induce the synthesis of HSPs, result in tolerance to much higher temperatures. The kinetics of HSP induction closely parallel the kinetics of thermotolerance induction, and the decay of HSPs from the cell closely parallels the decline in thermotolerance. Moreover, HSP functions are not limited to conditions of high temperature. They are induced by many other stresses, including anoxia and reoxygenation, heavy metal ions, ethanol, and inhibitors of respiration or oxidative phosphorylation. In general, exposure to one form of stress provides tolerance to another, suggesting that the proteins have broad protective roles. Studies of the heat-shock response are, therefore, of practical interest. (The development of crop plants with extended heat tolerance, for example, would be of enormous benefit to agricultural productivity.)

HSPs are of particular interest to human biology and medicine, for four reasons. 1) 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. 2) 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. 3) 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. 4) The proteins interact with and potentiate the function of many other proteins in the cell.

As important as these pragmatic concerns may be, most studies of the heat-shock response, including those in my laboratory, have been motivated by other considerations. Initially the response was exploited as a model system to study the cellular mechanisms involved in regulating protein synthesis. Because induction of the proteins is required for survival, a number of sophisticated 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, or their close relatives, play vital roles in the cell at normal temperatures has opened up a whole new field of investigation. The specific molecular functions of the HSPs are only beginning to be elucidated, but they appear to play a role in a remarkable number of basic cellular processes, among them secretion, signal transduction, and ribosome assembly. Determining the roles these proteins play in these processes will provide fundamental insights in cell biology.

We are investigating the regulation and the function of the HSPs. 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 them. 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 the synthesis of HSPs after heat shock. Tom McGarry and Bob Petersen have found that heat-shock mRNAs in *Drosophila* cells are preferentially translated during heat shock by virtue of sequences in their 5′-untranslated leaders. They are preferentially repressed after heat shock (during recovery) by virtue of sequences in their 3′-untranslated tails, sequences that selectively target them for destruction. The latter sequences are shared by many non-heat-shock 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 messenger RNA precursors, explaining the puzzling observation that heat-shock genes generally do not have intervening sequences. In this line of research our studies of HSP function and regulation overlap. Mild preheat treatments, which induce the synthesis of HSPs, protect RNA processing from disruption during heat shock. Examination of yeast cells carrying mutations in various *HSP* genes suggests that both the *HSP104* and the *HSP70* genes play a role in protecting processing at high temperatures.

To investigate the function of the HSPs, we have created a series of mutations in the genes of both yeast and *Drosophila*. Kathy Borkovich found that hsp83 is essential for growth at all temperatures in yeast cells, but that it is required at higher concentrations for growth at higher temperatures. Thus induction of this protein 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 to drive the equilibrium of these interactions toward complex formation. Recently, in collaboration with Keith Yamamoto's laboratory, we have demonstrated that hsp82 interacts with the steroid hormone family of receptors and helps these proteins fold into an active conformation.

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. More recently, we have found that this protein is highly conserved in mammals and in prokaryotic cells. Moreover, it appears to provide protection from many other forms of stress, such as exposure to ethanol and sodium arsenite.

In *Drosophila* our mutational analysis has concentrated on hsp70. Jan Rossi found that varying the level of hsp70 expression in *Drosophila* cells varies the rate at which 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. This particular protein thus helps to protect cells from the ravages of extreme temperatures but is disadvantageous at normal temperatures. Further experiments will focus on examining the specific cellular and developmental processes in which hsp70 plays a role.

Finally, Kent Golic has taken advantage of the heat-shock response to devise an inducible system for site-specific recombination in *Drosophila*. He placed the *FLP* recombinase gene of yeast under the control of the hsp70 regulatory sequences and transformed *Drosophila* embryos with this construct. He transformed other embryos with a recombinase target, consisting of an eye color gene flanked by target sequences for FLP recombination. When the two strains were crossed and their progeny were heat-shocked, the eye color gene was excised in both somatic and germline tissues, with the frequency of excision depending on the severity of the heat shock. This system is being used in our laboratory to vary the dosage of *hsp* mutants, but it has many other potential applications in *Drosophila* and other organisms.

T Cell Surface Glycoproteins in Development and Viral Infections

Dan R. Littman, M.D., Ph.D.—Assistant 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. After a postdoctoral fellowship in Richard Axel's laboratory at Columbia University, where he isolated the genes for CD4 and CD8, he assumed his present position at the University of California, San Francisco.

MATURATION of T lymphocytes involves the interaction of surface proteins on these cells with components of the thymic microenvironment. Thymocyte precursors (immature thymus-generated cells) bearing appropriately rearranged and assembled T cell receptors are selected through this process of cell-cell interaction to mature and migrate to peripheral lymphoid organs. Selection involves elimination of self-reactive T cells and preferential outgrowth of cells capable of reacting to foreign antigen complexed to major histocompatibility complex (MHC) molecules.

Aside from the T cell receptors, two other cell surface glycoproteins, CD4 and CD8, play central roles in this selection process. Both CD4 and CD8 molecules are expressed on immature thymocytes, but upon maturation the gene for one or the other is shut off. Cells that have receptors for class I MHC molecules continue to express CD8, but shut off CD4; cells with receptors for class II MHC 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 having appropriate specificity.

We have previously used a cell-cell adhesion assay to demonstrate that the CD8 molecule binds to the membrane-proximal domain of the class I molecule. Since the T cell receptor binds to another region of this molecule, it is likely that the MHC molecule serves as a bridge to bring the receptor complex and CD8 into close contact.

We have now tested the hypothesis that binding of CD8 to class I MHC during thymic differentiation is required for the maturation of class I–restricted T cells. Transgenic mice that express a mutant form of MHC-I that cannot bind CD8, but can interact with T cell receptors, were shown to be defective in developing a mature T cell repertoire. Analysis of T cells in these mice provides evidence that CD8-MHC binding is required both for intrathymic deletion of self-reactive T cells and for positive selection of useful T cells.

To study the roles of CD4 and CD8 in greater detail, we are preparing animals that are defective in the expression of these molecules. Namely, we have utilized gene-targeting technology to mutate the CD4 gene in embryonal stem cells and have injected these cells into mouse blastocysts, generating chimeric animals. We are attempting to propagate the mutant gene in the mouse germline, and the immune system of mice lacking CD4 expression will be studied. We have also prepared transgenic mice expressing mutant forms of CD4 that are predicted to be defective in signal-transducing functions, and these will be used to analyze CD4 function in the absence of endogenous CD4 expression.

The CD4 and CD8 glycoproteins have been shown to function importantly in the activation of peripheral T lymphocytes. For example, artificial crosslinking 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.

The CD4 glycoprotein is doubly important because it is the receptor for the human immunodeficiency virus (HIV). We have completed a muta-

genesis analysis to identify the region of CD4 that binds to the viral envelope glycoprotein. This information, coupled to the recent elucidation of CD4's three-dimensional structure, may facilitate the development of agents that block binding.

Although human cells expressing the CD4 molecule are susceptible to HIV infection, murine cells expressing CD4 bind virus but are resistant to infection. This resistance, we have found, is due to the inability of the viral envelope to fuse to the murine plasma membrane. It is likely that a human-specific plasma membrane molecule other than CD4 is required for fusion of virus to target cells. To identify such a molecule and to isolate its gene, we have packaged a selectable marker within the HIV particle. Such particles bind to CD4-bearing human cells, are internalized, and are processed to generate a proviral DNA that is integrated in the host genome. The cells are then grown in selection media.

These virus particles do not infect CD4-bearing murine cells. But if the particles are prepared with an envelope glycoprotein that can normally interact with murine cells, the selectable HIV genome is integrated in the cells. Efforts are under way to transfer human genes into the murine cells to permit cell survival after infection with the selectable virus. It is hoped that identification of a second molecule involved in HIV entry may lead to the design of novel agents to interfere with the virus's spread.

We have utilized this system to demonstrate that the HIV envelope glycoprotein can be replaced by that of the other pathogenic human retrovirus, human T cell leukemia virus (HTLV), to form 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. Thus 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.

Antigen-Specific T Cell Receptors

Dennis Y.-D. Loh, M.D.—Associate Investigator

Dr. Loh is also Professor of Medicine, Genetics, and Microbiology at Washington University School of Medicine 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. He took a clinical residency in internal medicine at the Peter Bent Brigham Hospital, Boston, and then 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. Its critical role in maintaining health is well typified in the acquired immune deficiency syndrome (AIDS), in which destruction of a specific portion of the immune system results in a fatal disease.

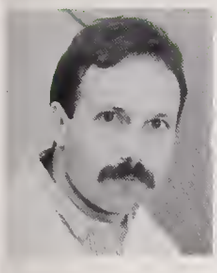
Our investigation is focused on the molecular mechanisms that allow normal and abnormal development of the antigen-specific T cells. T cells are those lymphocytes (a type of white cell) that depend on the presence of the thymus gland for maturation. An antigen may be viewed as any marker that these cells 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.

During the past few years, my laboratory has concentrated on identifying the genetic elements that encode the TCR genes. Recently we have shifted our efforts to study the function of T cells in the intact animal. 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 were not physically eliminated but are functionally paralyzed. These studies allow us to lay the foundation to study how T cells acquire self-tolerance.

Most recently, we have developed a new generation of TCR transgenic mice in which we can induce massive cell death when the appropriate antigen is administered *in vivo*. These studies prove that clonal elimination of self-reactive cells is indeed mediated by programmed cell death. These new approaches should shed insight into the development of self-tolerance and the mechanism of programmed cell death during development. Since distinguishing what is self and nonself is a central problem in immunology, we hope these studies will increase understanding of transplantation rejection and autoimmune phenomena.

Molecular Genetics of Mammalian Glycosyltransferases



John B. Lowe, M.D.—Assistant Investigator

Dr. Lowe is also Assistant 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. He was trained in clinical pathology and molecular genetics at Washington University School of Medicine in 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 molecules called oligosaccharides that are found on the surface of mammalian cells and to explain how the cells regulate expression of these molecules. Oligosaccharides consist of many different 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 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 the processes that allow mammalian cells to 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 have focused on developing systems to isolate glycosyltransferase genes without the benefit of purified enzyme protein. In one instance, we isolated the gene for human H blood group $\alpha(1,2)$fucosyltransferase, using a scheme in which the gene was first transferred into a cultured mouse cell line by a process called transfection. We were then able to use molecular cloning procedures to retrieve the human gene from the mouse genome.

We have also used a similar technical approach to isolate other glycosyltransferase genes. This scheme involves transfer of cloned segments of human genes back and forth between animal cells and the bacterial host cells used for molecular cloning purposes. With this approach an $\alpha(1,3/1,4)$fucosyltransferase gene, for example, was isolated that represents the human Lewis blood group locus. We have recently used the H and Lewis gene segments and cross-hybridization procedures to isolate still other glycosyltransferase genes with related or unique structural and functional properties.

The cloned gene segments in each case represent tools for investigating the function and regulation of cell surface oligosaccharides. For example, we recently used these 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 neutro-

phils 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 endothelial leukocyte adhesion molecule I, or ELAM-I, that is found on the surface of activated endothelial cells. Structural features exhibited by ELAM-I suggested that it might interact with an oligosaccharide molecule specific to the surface of neutrophils, but this hypothesis had not been confirmed, nor had the nature of such molecules been defined. By transfecting different glycosyltransferase gene segments into cultured mammalian host cells, we were able to recapitulate the biosynthesis of several distinct sets of cell surface oligosaccharide molecules. We found that one set allowed transfected cells to adhere to ELAM-I.

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. Subsequent work by our own group and others has confirmed that these and other glycosyltransferase genes play important regulatory roles in the biosynthesis of oligosaccharides recognized by ELAM-I.

Recently it has also been shown that oligosaccharides related or identical to the sialyl Lewis X determinant may function in adhesive interactions exhibited by two other cell surface proteins, called GMP-140 and Mel14/Leu8. These latter proteins, structurally related to ELAM-I, also function to mediate leukocyte-endothelial cell adhesion in inflammation and lymphocyte recirculation.

Circumstantial evidence gathered by other investigators suggests that oligosaccharides are importantly involved in cell adhesion during mammalian embryogenesis. We are directing current 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. 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.

AN emerging theme in vertebrate development is that temporally and spatially regulated signals between cell populations direct pattern formation. The process by which such signals effect differentiation of a tissue is called induction. A major goal in our laboratory is to understand what these inductive signals are and how their activities are orchestrated.

Recently two major classes of genes have been found to have an important role in vertebrate development. Homeobox genes, the first of these classes, control the identity of individual body parts. They have in common a 180–base pair sequence element, the homeodomain. This genetic element, highly conserved in evolution, has been shown to confer the sequence-specific binding of the homeoprotein to DNA, with the remainder of the protein functioning as a transcriptional activator or repressor. Moreover, in the examples analyzed, the homeodomain alone can interact with DNA. Although the target genes for homeobox proteins are known in some cases, particularly in *Drosophila*, and include other homeotic genes, they are not well identified or understood in mammals.

From work in both *Xenopus* and *Drosophila*, it has become clear that certain growth factors constitute a second class of gene products important in morphogenesis. One hypothesis is that individual homeobox genes, acting alone or in concert, also regulate, perhaps directly, the function of individual growth-controlling peptides. It is already known that members of a family of peptide growth factors, called transforming growth factor-βs (TGF-βs), can influence the expression of certain homeobox genes. A possibility that individual homeobox genes might also regulate TGF-βs or other growth factors is suggested by coincidental patterns of expression in developing vertebrate embryos, as well as from genetic experiments in *Drosophila*. This raises the general question, What are the transcriptional targets of individual homeobox-containing genes during vertebrate embryogenesis? We have taken a multidisciplinary approach to this question, with both general and specific avenues of investigation.

Homeobox Genes in the Developing Mouse Heart

The embryonic development of the vertebrate heart consists of an interaction between two concentric layers of epithelium—an outer epithelial layer, or myocardium, and an inner epithelial layer, or endocardium. The outer epithelial layer has been shown to elaborate a soluble factor that, in concert with TGF-β1, triggers the transformation of the inner epithelium into an undifferentiated mesenchyme. This subsequently differentiates into the connective tissue of the valves of the adult heart.

In the mouse embryo, TGF-β1 is expressed in the atrioventricular canal, the site of valve formation, at the time of differentiation. Interestingly, a homeobox gene of the *msh* class, *Hox-7.1*, appears to share during development a common pattern of expression with TGF-β1 in the atrioventricular canal. We have cloned the TGF-β1 promoter and regulatory region and are testing its ability to interact with recombinant *Hox-7.1* homeobox proteins. We also plan to determine the sequence of DNA responsible for binding *Hox-7.1*, with a view to comparing the binding sites of *msh* class homeoboxes with those of the *Antennapedia* class, which are already known.

Using a degenerate set of primers engineered to the most highly conserved regions of the homeobox, we have also amplified and analyzed additional homeobox sequences from reverse-transcribed embryonic cardiac RNA from day 10 of mouse embryogenesis.

Homeobox Gene Expression in Developing Embryonic Mouse Kidney

A second system studied in our laboratory is the developing metanephros, or kidney. Approximately 20 different homeobox genes have been identified in the day-15 kidney of the embryonic mouse. These genes have been divided into three major groups: 1) genes identical to known homeobox sequences, 2) genes that are clearly different from any known sequences, and 3) genes that are very closely related to known sequences, but still different. Immediate goals for the ho-

meobox sequences that are clearly novel are to obtain cDNA clones, to map them in the mouse genome, and to investigate their expression patterns by *in situ* hybridization in the embryonic kidney. A longer-term goal is to identify the target genes that these homeoboxes interact with in mouse kidney development.

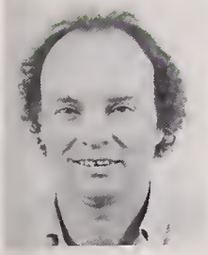
Approaches to Identifying Target Genes of Homeobox Proteins

Since vertebrate development appears to proceed by a sequential hierarchy of transcriptional activation (and repression), and since homeobox genes are apparently important in mediating such events, we have sought to develop a technique for identifying genes that homeobox genes directly regulate. One technique being explored makes use of a yeast expression system in which an individual yeast cell contains an expression plasmid for the homeoprotein whose binding site is being sought, fused to Gal4, a strong transcriptional activator. This homeobox-Gal4 fusion protein is placed under the control of an inducible promoter, so that the synthesis of the protein can be turned on or off.

In yeast, such a system has been shown capable of activating a target plasmid containing a sequence-specific binding site upstream of a promoter controlling the expression of the bacterial gene *lacZ*, which thus serves as a reporter gene. Conveniently, *lacZ* encodes an enzyme, β-galactosidase, that will cleave a fluorescent substrate, resulting in a fluorescent yeast cell. Such cells can be separated with the fluorescence-activated cell sorter (FACS). We hope to use this system to remove background from subgenomic libraries and then, after induction of homeobox expression, to sort cells containing specific homeobox binding sites.

A second general approach under consideration is the addition of labeled or retrievable homeodomains to chromatin, isolated from the developmental stage and tissue of interest. This is patterned after an analogous approach used in *Drosophila*. However, a current limitation is that high-titer monoclonal antibodies are not available for many mouse homeobox genes.

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, cells have evolved rigorous controls to ensure that both processes 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 decision to get ready for DNA synthesis is made in the G_1 period (after mitosis but before DNA synthesis or S phase), and the decision to begin cell division (M phase) is made in G_2 (after S phase is complete). 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 malignancy.

$G_2 \rightarrow$ M Regulation

In the last three years great strides have been made in understanding the checkpoint governing entry into mitosis. For many years it had been known that mitosis is dominant over interphase. That is, the signal that tells a cell to go into mitosis, if inappropriately expressed earlier in the cycle, can cause premature mitosis and cell death. Our laboratory developed an interphase cell extract from frog eggs in which addition of mitotic signals caused synthetic nuclei in the extract to enter mitosis *in vitro*. We then purified the mitosis-signaling enzyme (called M-phase 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 subunit of the enzyme that catalyzes mitosis 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 phase, and are then degraded near the metaphase/anaphase transition in mitosis. This degradation is required 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. To study the differences in A- and B-type cyclins, we have expressed the frog cyclin genes in insect cells. This allows us to make large quantities of functionally active frog cyclins, which are able to drive synthetic nuclei into mitosis in extracts after complexing with *cdc2* kinase.

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 is required for activation of MPF. 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 they 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 that a substrate for phosphorylation by *mos* exists that can lead to activation of MPF as well as stabilization of cyclin in meiosis II. This year we discovered that B-type *Xenopus* cyclins were *mos* phosphorylation substrates and that the affected sites were distinct from those phosphorylated by *cdc2* kinase itself. As soon as the exact site of *mos* phosphorylation in cyclin is determined, we can ablate the site(s) and evaluate effects on cyclin function and cell cycle control.

Reentry of Quiescent Cells Into the Cell Cycle

A vast fraction of the body's cells do not actively participate in the cell cycle. Instead, they are in a quiescent state awaiting a signal to reenter the cycle. This signal is often in the form of a small protein termed a growth factor. Some forms of cancer appear to arise because cells reenter the cycle in the absence of a growth factor, leading to uncontrolled cell growth. Some cells enter quiescence in the G_2 phase. However, most quiescent cells waiting for proliferative signals are in a specialized state of the G_1 phase of the cell cycle known as G_0. Many normal cells often enter this state when deprived of growth factors. It is evident that the transition into and out of the G_0 state is related to the general question of how a cell decides it is time to prepare for DNA replication. Available evidence suggests that once a cell is committed to replicate its DNA, it almost invariably progresses on through the cell cycle until the $G_2 \rightarrow M$ checkpoint controls become activated.

Many of the growth factors stimulate quiescent cells to reenter the cell cycle by binding to specific receptors on the cell surface. These receptors are actually enzymes expressing a protein kinase activity that transfers phosphate groups to tyrosine residues on substrate proteins. Many oncogenes that cause cancer turn out to be mutant versions of these growth factor tyrosine kinases that signal the cell to reenter the cycle in an uncontrolled fashion. We have undertaken an analysis of the steps in signaling by tyrosine kinases, termed the signal transduction pathway, in order to gain a better understanding of how to control inappropriate signaling. Our approach has been to characterize biochemically an event rapidly stimulated by activated tyrosine kinase receptors and then work backward up the signaling pathway to define the signal transduction mechanism.

Over the past five years we have identified in ribosomes a protein called S6 that is phosphorylated on serine residues very rapidly after activation of tyrosine kinase receptors. This means the signal transduction pathway involves receptor tyrosine kinase activation of a serine kinase. We have shown that the activation does not occur by direct phosphorylation of the S6 kinase on tyrosine residues, indicating that there are intermediary molecules that participate in the signaling mechanism. Instead, the S6 kinase is activated by direct phosphorylation on serine and threonine residues by another serine/threonine protein kinase called microtubule-associated protein 2 kinase or MAP kinase. *Xenopus* MAP kinase has been purified this year and found to be itself phosphorylated on both threonine and tyrosine residues by additional upstream protein kinases. This means we are getting very close to the direct tyrosine phosphorylation events stimulated by a tyrosine kinase receptor.

In both G_1 phase after the G_0 transition and in G_2-arrested cells, *cdc2* kinase becomes activated after reentry into the cell cycle and may also experience an increase in its rate of synthesis, indicating that it works at both checkpoints in the cycle. In the G_1 phase, *cdc2* kinase is complexed with a different group of proteins, known as G_1 cyclins, whose synthesis increases in late G_1. A full understanding of cell cycle control will come when the entire sequence of events from tyrosine kinase activation to *cdc2* and G_1 cyclin synthesis is elucidated.

The Role of T Cells in Health and Sickness

Philippa Marrack, Ph.D.—Investigator

Dr. Marrack is also a member of the Department of Medicine at the National Jewish Center for Immunology and Respiratory Medicine and Professor of Biochemistry, Biophysics and Genetics, of Microbiology and Immunology, and of Medicine at the University of Colorado Health Sciences Center, Denver. She was educated at the University of Cambridge, England, and took her Ph.D. in biological sciences at the MRC Laboratory for Molecular Biology in Cambridge. She came to the United States to do postdoctoral work with Richard Dutton at the University of California, San Diego. From there she moved to the University of Rochester and then, after seven years, to her present position. Dr. Marrack was awarded the Wellcome Foundation Prize by the Royal Society and is a member of the National Academy of Sciences.

T cells are absolutely essential to the health of higher vertebrates. They patrol the body, recognize some invaders such as bacteria and viruses, and orchestrate their destruction. The importance of these cells is tragically revealed by the devastating effects of AIDS, in which a certain type of T cell is inexorably destroyed.

T cells recognize targets, using receptors on their surfaces. Each T cell has about 20,000 identical receptors, and those on any given T cell differ from those on other T cells. Thus a human being has about a million million T cells, all bearing different receptors to help in defense. When an organism invades the body, a wide array of T cell receptors is available to bind to fragments of the organism, called antigens. By chance the receptors on some of the T cells will be able to bind to these antigens. These T cells bearing bound receptors are stimulated to multiply and destroy.

How does the body develop so many different T cell receptors? This is achieved by combining in the receptor one each of several elements for engagement. T cell receptors are made up of five variable elements: Vα, Jα, Vβ, Dβ, and Jβ. Each T cell picks for its receptor 1 Vα out of about 50 available Vα's. Likewise the cell chooses 1 Jα from a collection of about 50 Jα's and 1 Vβ from about 50 Vβ's. Dβ and Jβ are chosen from pools of about 6 Dβ's and 13 Jβ's, respectively. By combining these different components in different ways, the total T cell pool manages to express many receptors and hence is able to bind many foreign organisms.

One of the major theoretical problems in immunology has always been the matter of tolerance to self. If our T cells can recognize and destroy most foreign material—including, for example, skin grafts from some close relatives—how is it that our T cells do not attack and destroy *us*? Recently we have shown that tolerance to self is created by at least two different means. The first involves T cell death. As T cells are developing they go through a stage at which the binding of material to their receptors causes them to die. Developing T cells are almost bound to be exposed to self; therefore cells that are potentially self-reactive will die at this stage. At a later stage in the life history of T cells, another mechanism comes into play, and T cells that can react with self appear to be inactivated.

We do not yet understand why T cells that can react with self are killed in some cases and inactivated in others. More importantly, perhaps, we do not understand why T cells that can react with self sometimes escape both these processes and emerge later to cause autoimmune diseases. There is good evidence, for example, that juvenile diabetes occurs because T cells in some children are able to bind to material made by the pancreas. The T cells in question exercise this ability and destroy the pancreatic cells that secrete insulin.

In order to study this problem, we have recently started to work on T cells in humans, in collaboration with Brian Kotzin, a rheumatologist at the National Jewish Center. Even though T cell receptors comprise five variable components, Vα, Vβ, etc., we found to our surprise that in mice we could understand a lot about T cell specificity, responses, and autoimmunity by studying just one part of the T cell receptor, Vβ. For this reason we have begun our clinical studies by developing tools to explore Vβ use by T cells.

At first we developed a molecular biological method—quantitative polymerase chain reaction—to measure what proportion of T cells in different individuals bore a particular Vβ. Using this method, we showed that patients with toxic shock syndrome have very large numbers of T cells bearing Vβ2. This occurs because the bacteria that cause toxic shock, *Staphylococcus aureus*, produce a toxin that stimulates human T cells bearing Vβ2 to divide. We believe, in fact, that the disease toxic shock is caused by massive replication of Vβ2-bearing cells and their copious secretion of lymphokines, hormone-like materials that are useful to the body in small quantities but which at high concentrations cause shock and death.

At the moment we are applying these same methods to another human disease, rheumatoid arthritis. Although the reasons some persons develop this disease are not fully understood, there are many indications that rheumatoid arthritis is triggered by a T cell malfunction. We have examined T cells from the rheumatic joints and blood of patients with this disease and have found that T cells bearing Vβ14 are selectively deficient in the blood. Moreover, there are indications that such T cells may accumulate in the diseased joints. Now we are testing the ideas that rheumatoid arthritis is caused by malfunction in T cells bearing this Vβ and that it may have been started by a toxin-like substance that stimulates T cells bearing Vβ14.

The quantitative polymerase chain reaction used in these experiments is a cumbersome and difficult technique. With a view to making experiments easier, we are developing a collection of monoclonal antibodies specific for different human T cell Vβs.

Control of Cell Growth and Phenotype by Transforming Growth Factors

Joan Massagué, Ph.D.—Investigator

Dr. Massagué is also 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. His post-doctoral training was with Michael Czech on the identification of insulin-like growth factor receptors. He was Assistant and Associate Professor of Biochemistry at the University of Massachusetts Medical School (Worcester) before moving to his present position.

THE purpose of our research is to elucidate the mechanism of action of transforming growth factor-β (TGF-β) and other factors. Our appreciation of the complexity of the TGF-β system has escalated dramatically in the last five years. TGF-β is now known to represent a vast family of paracrine polypeptides thought to be involved in many processes of tissue development, morphogenesis, recycling, and repair. This family includes homodimeric and heterodimeric products of more than 20 genes. The distribution of TGF-β-related factors is widespread in organisms from fruit flies to humans, and their evolutionary conservation is unusually strict. They are actively expressed throughout embryo development and into adulthood, in discrete regions of many tissues with characteristic temporal patterns.

TGF-β exists in multiple isoforms, including TGF-β1, -β2, and -β3 in mammals. These isoforms show many similarities in their biological properties, but also some important differences. TGF-β is exceptional in its multifunctional nature, acting, for example, as a paracrine inhibitor of cell proliferation and a regulator of cell differentiation, cell adhesion, and extracellular matrix deposition. Defects in TGF-β activity as a growth suppressor may lead to oncogenesis, and its excessive activity to fibrotic disorders.

Mechanisms of TGF-β Action

One of our goals is to identify, clone, and characterize the membrane receptors that mediate TGF-β action. To this end, we have identified membrane proteins that bind TGF-β with high affinity. These include TGF-β receptor components I and II, which are distinct cell-surface glycoproteins of 53 and 70 kDa, respectively. They are expressed ubiquitously, but in low numbers, in most normal and transformed cells, with the notable exception of human retinoblastoma cells. These receptors markedly discriminate between the various TGF-β isoforms. A third TGF-β-binding component, betaglycan, has an unusual structure; it is a membrane-anchored proteoglycan.

Which of these binding components mediates TGF-β action? The resolution of this question made use of one of the most remarkable effects of TGF-β, inhibition of cell growth. A panel of mutant cell clones resistant to the growth inhibitory action of TGF-β was isolated by selection of chemically mutagenized cells that would grow in the continuous presence of TGF-β. Characterization of these mutants showed that the mutations affected the expression of TGF-β receptors I and II but not betaglycan. Moreover, somatic cell hybrids between cell mutant clones demonstrated that all TGF-β responses were regained when expression of receptors I and II was restored by complementation between different cell mutants. This evidence linked receptors I and II to mediation of the pleiotropic action of TGF-β and indicated that both receptor components cooperate to mediate the multiple effects of TGF-β. We are developing methodology to use the TGF-β receptor-defective cells to clone receptor cDNAs.

Progress has also been made in the study of betaglycan. This integral membrane proteoglycan contains heparan/chondroitin sulfate glycosaminoglycan (GAG) chains attached to a 110- to 130-kDa core glycoprotein. The GAGs are not required for TGF-β binding or expression of betaglycan on the cell surface; the TGF-β-binding site resides in the core protein. Soluble forms of betaglycan lacking membrane anchor are released into the medium and can be found in extracellular matrices.

Betaglycan is intriguing as a novel example of regulatory molecules involved in the biology of growth factors. It is widely distributed in tissues, but absent from certain cell types that, nevertheless, respond to TGF-β. No evidence is presently available for a direct role of betaglycan in signal transduction. It may be involved, however, in ligand presentation to TGF-β receptors I and II and could function as a pericellular reservoir of bioactive TGF-β. Like other membrane proteoglycans, betaglycan may also participate in cell adhesion and recognition. We have purified it to homogeneity in order to investigate it further.

Recent progress linked growth arrest by TGF-β in late G1 to its ability to control the phosphorylation state of an intracellular growth suppressor,

RB, the nuclear phosphoprotein product of the retinoblastoma susceptibility gene. TGF-β arrests cell growth at precisely the cell cycle point where RB phosphorylation is due to take place. This observation led to the finding that TGF-β inhibits RB phosphorylation, retaining this protein in its presumed growth suppressor state.

Expression of the SV40 transforming protein, large T antigen, which binds unphosphorylated RB and disrupts its growth-suppressive function, prevented growth inhibition by TGF-β without affecting its inhibition of RB phosphorylation or other effects. Thus TGF-β's inhibition of RB phosphorylation is not an irrelevant event but, possibly, one that mediates cell cycle arrest by this factor. These results represent a case in which a prototypic intracellular growth suppressor (RB) participates in the mechanism of action of a growth-inhibitory paracrine agent (TGF-β).

The mechanisms involved in growth suppression by TGF-β might also relate to those that lead to its control of cell differentiation. This effect was best studied with skeletal muscle myoblasts. TGF-β inhibited myoblast differentiation in part by interfering with the expression of myogenin, a differentiation-determining gene. Other studies showed that TGF-β is a major regulator of the cell adhesion apparatus—cell adhesion receptors and extracellular matrix protein expression—in many cell types. These effects could in turn explain the ability of TGF-β to affect the morphology and phenotype of cells and, most importantly, explain its presumed roles in tissue morphogenesis and fibrotic disorders. Effects of TGF-β on gene expression might be mediated by its ability to control the levels or activity of certain transcription factors, a possibility that is the focus of current studies.

Cell-Cell Communication by Membrane-anchored Growth Factors

Our research interests extend also to the area of growth factors that are synthesized as part of transmembrane protein precursors. The model system in our studies is transforming growth factor-α (TGF-α). This is a 50–amino acid polypeptide initially identified in culture fluids of transformed cells, hence its name. It is not related to TGF-β but shares structural homology with members of the epidermal growth factor (EGF) family. TGF-α is most prevalent in tumor-derived cells and causes a high incidence of mammary and liver neoplasias in transgenic mice that chronically overexpress it. The factor stimulates cell growth through the EGF membrane receptor, which is a ligand-activated tyrosine kinase.

Paracrine growth factors and polypeptide hormones are generally synthesized as larger soluble precursors that are cleaved by specific peptidases to release the bioactive domain. In contrast to this, a remarkable feature of TGF-α, shared with the other members of its family, is its generation by cleavage of a membrane-anchored precursor, pro-TGF-α. This is a 160–amino acid polypeptide with the TGF-α sequence in the extracellular domain. The proteolytic process that releases mature TGF-α from its precursor is inefficient in most cell types. Consequently, substantial levels of pro-TGF-α can accumulate on the cell surface.

Membrane pro-TGF-α can bind EGF receptors on the surface of adjacent cells. This interaction leads to signal transduction and DNA replication. Pro-TGF-α also functions as a mediator of cell-cell adhesion. Thus a membrane-anchored growth factor and its receptor can function simultaneously as mediators of cell-cell adhesion and as initiators of mitogenic stimulation by cell-cell contact. The membrane-anchored forms could be important in tissue development processes whose guidance depends on discrete cell-cell interactions incompatible with the diffusible nature of soluble factors. Given their structure, the membrane forms might function as signaling receptors. This possibility is under study.

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 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 proteins 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 at random with little apparent effect on folding or stability. These amino acids seem to be unimportant in protein folding. 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.

To try to simplify the complexity of the folding problem, an attempt has been made to replace some of the "nonessential" amino acids with alanine in phage T4 lysozyme. Such a "polyalanine" protein would, in principle, truncate all nonessential side chains and permit focus on those parts of the amino acid sequence that are critical for the folding process.

As a first step, a series of alanines was introduced within the α-helix that includes residues 126–134 of T4 lysozyme. The somewhat surprising result was that alanines were not only tolerated at most positions in the α-helix; they actually increased the protein's stability. This indicates that alanine is a strongly helix-favoring residue. It also suggests that the replacement of solvent-exposed residues of α-helices with alanines might provide a means of increasing the stability of other proteins. Finally, the fact that a series of alanines can be introduced into T4 lysozyme confirms that this might be a way to simplify the protein folding problem.

Engineering Proteins of Enhanced Stability

We are using the lysozyme from bacteriophage T4 to define the contributions that different types of interaction (hydrogen bonds, hydrophobic interactions, salt bridges, etc.) make to the stability of proteins. Much of our emphasis during the past year has been on gaining a better understanding of electrostatic interactions. It is known that they can be strong in some cases but seemingly weak in others. To try to explain this, we first investigated the effect of long-range electrostatic interactions by using genetic engineering to change the charge of a number of groups on the surface of phage T4 lysozyme. We were able to reduce the overall charge on the protein from +9 units to +1 unit. Nevertheless, there was almost no change in protein stability, indicating that the interaction between the different charged groups is very weak.

Next we examined short-range electrostatic interactions by introducing negatively charged groups adjacent to positively charged amino acids on the surface of the protein. Again, there was almost no change in the protein stability.

Finally, we introduced charged groups next to the ends of α-helices. In this case an increase in stability was consistently observed, indicating that there are favorable electrostatic interactions between charged amino acids and the charges on the ends of the α-helices. This is the so-called "helix dipole" effect.

It is at first surprising that electrostatic interactions between charged groups on the surface of a protein tend to be weak, whereas interactions with the α-helix dipole tend to be much stronger. The reason for the difference, we believe, is that charged groups on the surface of a protein tend to

be very mobile, whereas the α-helix dipoles are localized during the folding of the protein.

The effectiveness of the α-helix dipole in contributing to protein stability is a verification of our earlier results. Such studies are indicating ways in which the stabilities of proteins might be improved by genetic engineering. We hope to be able eventually to stabilize biological molecules used in medicine, such as vaccines, and to improve enzymes that are used in the pharmaceutical and food industries.

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 was found to be very similar to that of interleukin-1β. It seems clear that 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.

Protein-DNA Interaction

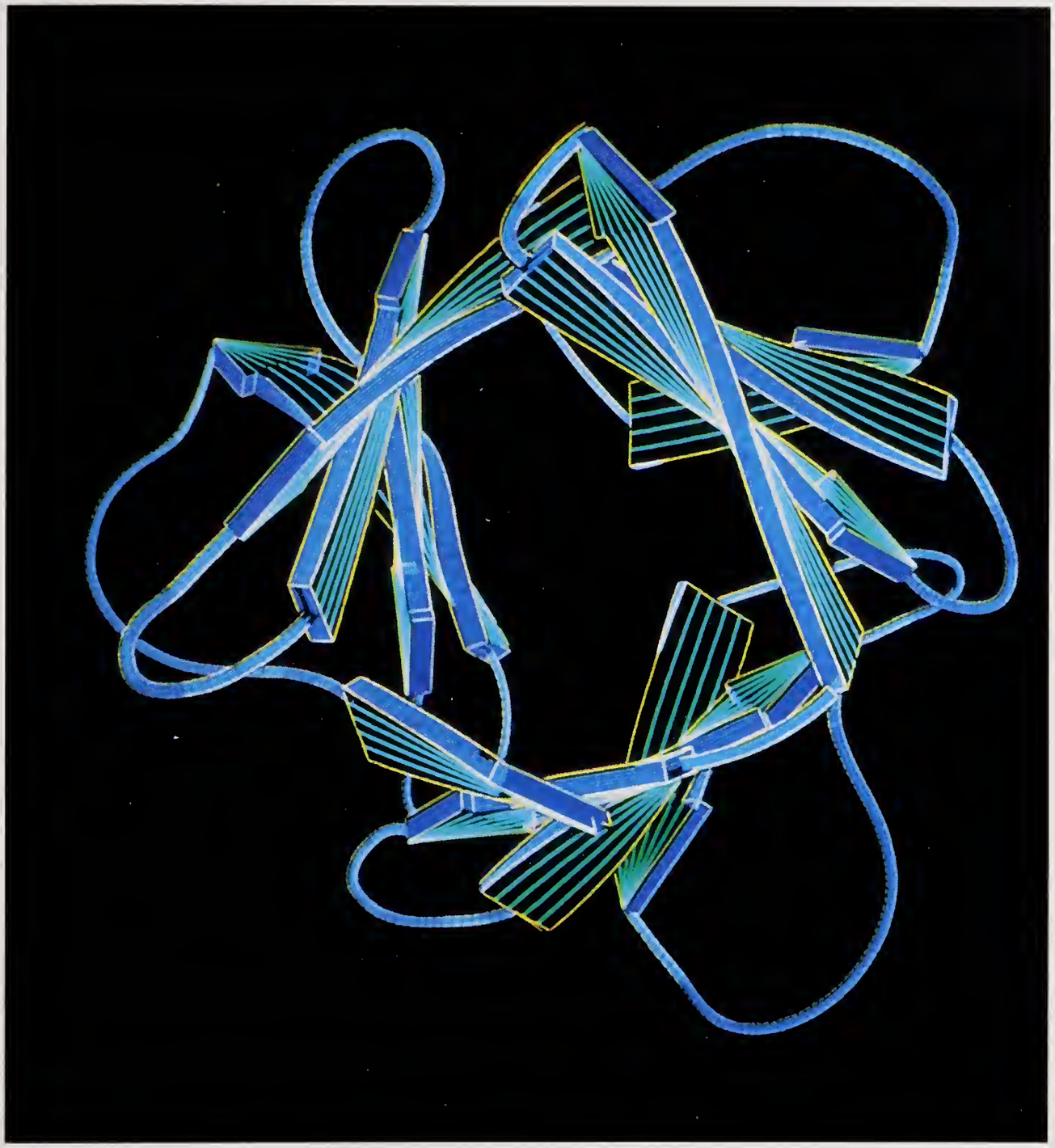
We have been interested for some time in the interaction between proteins and nucleic acids. In 1981 we determined the structure of the Cro repressor protein of bacteriophage λ (bacteria-infecting virus). Cro has served as one of the prototypical examples of a DNA-interacting protein. This small dimeric protein recognizes and binds to a specific set of 17 base pairs on the phage genome.

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. It 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.

Recently we have determined the crystal structure of Cro protein in complex with a tight-binding, 17–base pair DNA operator. We are improving the accuracy of the structure by a process of crystallographic refinement. 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.

The Cro dimer, however, undergoes a substantial conformational change relative to the uncomplexed crystal structure. One monomer rotates relative to the other by about 45°. This supports the idea that the Cro dimer is very flexible in solution. The DNA maintains an essentially Watson-Crick conformation, but is bent in the middle of the operator into the shape of a shallow boomerang.

We have recently obtained high-quality crystals of the biotin repressor from *Escherichia coli*. Determination of the structure of this more complicated protein, which not only binds DNA but also acts as an enzyme, is well under way.



Structure of fibroblast growth factor. Figure is a stylized computer-graphics representation of the three-dimensional structure of human fibroblast growth factor. The core of the molecule, consisting of a series of 12 extended segments (shown as arrows), appears to be common to a number of growth factors. The loop-like connections on the surface include the receptor-binding region and the heparin-binding region, which are seen to be far apart.

Research and photograph by Elizabeth Eriksson and Larry Weaver in the laboratory of Brian W. Matthews.

Transcription Factor Interactions

Steven Lanier McKnight, Ph.D.—Investigator

Dr. McKnight is also a staff member of the Department of Embryology in the Carnegie Institution of Washington, Baltimore, and an adjunct faculty member in the Departments of Biology and of Molecular Genetics at the Johns Hopkins University. 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.

FOR the past several years my colleagues and I have studied a mammalian transcription factor termed CCAAT/enhancer-binding protein (C/EBP). This protein is capable of binding to DNA in a sequence-specific manner and thereby regulating gene expression. By studying the detailed properties of C/EBP, we have sought to develop a better understanding of how genes are regulated in mammalian cells.

These efforts have been rewarded by a surprising discovery. The mechanism by which C/EBP binds DNA is common to that used by many other sequence-specific DNA-binding proteins, including the products of two important proto-oncogenes. The observations that led to this discovery were as follows.

After sequencing the gene encoding C/EBP, we fed its conceptually translated amino acid sequence into a computer database of known protein sequences. Much to our delight, a 60–amino acid segment of C/EBP proved to be sequence related to the products of the *Fos* and *Jun* proto-oncogenes. Earlier studies on C/EBP had shown that this same region of the protein was responsible for its ability to bind DNA. In other words, the region of C/EBP that was related in amino acid sequence to the FOS and JUN proteins was its DNA-binding domain.

Next we began to focus attention on the amino acids that evolution has conserved among the three proteins. We reasoned that the related sequences might reveal the underpinnings of a structural motif that facilitates specific interaction with DNA. The amino acid sequences shared by the three proteins occurred in two patches. One patch contained a conserved set of basic amino acids—arginines and lysines. On the carboxyl-terminal side of this "basic region," each protein exhibited leucine residues spaced every seven amino acids.

Bill Landschulz, then an M.D./Ph.D. student in the laboratory, noticed that neither the basic region nor leucine repeat region contained proline or glycine residues. Since prolines and glycines tend to be incompatible with α-helical structure, we reasoned that much of the DNA-binding domain of C/EBP, as well as FOS and JUN, might be α-helical. This prediction of α-helical structure provided an attractive role for the repeating leucine residues. Since the repeating period of α-helices is 3.5 amino acids per helical turn, and since the leucines were spaced at a heptad interval, the putative helix projected a continuous array of leucine residues from one helical face.

What might be the role of an α-helix that displays repeating leucines along one of its sides? Chemists have long known that leucines are unusually hydrophobic. Rather than being exposed to an aqueous or hydrophilic environment, hydrophobic amino acids prefer to interact with other hydrophobic compounds, most often within the internal fold of a protein. Following this lead we hypothesized that the α-helical region displaying a repeated array of leucines might represent a dimerization interface, allowing two polypeptide chains to coalesce along the helical face that contained the repeating array of leucines. We termed this hypothetical structure the "leucine zipper."

Elegant experiments by Peter Kim (HHMI) and his colleagues at the Whitehead Institute confirmed the general tenets of the zipper hypothesis and further established that zippered helices associate with each other in a parallel orientation.

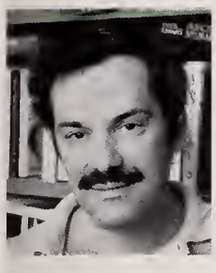
During the time that we were developing the zipper idea, research from several other laboratories demonstrated that the FOS and JUN polypeptides were capable of stable association. Recognizing this fact, Bill Landschulz, Peter Johnson, and I predicted that FOS and JUN would associate as dimers by virtue of their respective leucine zippers. Supportive evidence has been obtained by Robert Tjian (HHMI, University of California, Berkeley), Edward Ziff (HHMI, New York University Medical School), and Daniel Nathans (HHMI, the Johns Hopkins University Medical School).

If the leucine zipper region of each of these proteins is indeed responsible for allowing protein dimers to form, what is the role of the basic region? Knowing that the substrate for binding by these proteins is DNA—a negatively charged polymer—we reasoned that the basic region might facilitate direct contact with DNA. Such

speculation provided a satisfying fit between the dimeric proteins and their DNA-binding sites.

A common property shared by the binding sites for leucine zipper proteins is dyad symmetry. Two short sequences, related to each other as are the symmetric halves of a restriction endonuclease recognition site, are directly abutted. The basic region of one polypeptide chain would fit one half of the dyad symmetric binding site, while that of the other chain would fit the other half. Evidence has recently been obtained by a number of different laboratories in support of this idea. The acid test, however, will come upon resolution of the actual molecular structures of leucine zipper proteins. Paul Sigler (HHMI, Yale University) and Stephen Harrison (HHMI, Harvard University) are well on the way to crystallizing two different leucine zipper proteins.

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 ion fluxes giving rise to the generation, propagation, and integration of electrical signals in neurons, muscle, and other electrically active cells. By forming aqueous pores right through the heart of the channel protein (and hence across the membrane that the protein spans), channels act as "leakage" pathways for ions down their preestablished thermodynamic gradients. These proteins are 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 open and close their conduction pores rapidly in response to external signals, such as binding of neurotransmitters or changes in the transmembrane electric field.

Work here is directed toward questions of basic molecular mechanisms of ion channel operation and of the underlying protein structures involved. Since no high-resolution structures have been obtained for this class of proteins (and since none is coming over the horizon), one must draw structural inferences from close examination of ion channel function. This can be done because ion channels, unique among all classes of proteins, can be studied 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 the pore 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 recent advances in 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 up the channel's externally facing "mouth," we are now utilizing the peptide as a probe of this important region of the channel. Employing site-directed mutagenesis with the *Drosophila* Shaker K^+ channel, we have identified residues that locally and specifically alter the binding of CTX. These residues are evidently located near the ion entryway, so we are homing in on regions of the protein that form the transmembrane pore. Two developments of the past year have placed us in position to use this toxin as a structural probe of the K^+ channel's outer mouth. First, using two-dimensional nuclear magnetic resonance, we determined the solution structure of the toxin. Second, we constructed a synthetic gene for CTX and successfully overexpressed the fully functional peptide in *Escherichia coli*. Using high-level expression, routine structure determination of the toxin, and site-specific mutagenesis of both toxin and channel, we are now attempting to map the locations in the K^+ channel mouth of residues that make direct contact with residues on CTX.

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 solubilized state and are presently using it to perform conventional purification studies. We intend to study the purified channel protein as a supplement to structure-function work at the cDNA level on this recently cloned anion channel.

Structure-Function Relations in a Minimal K^+ Channel

We are beginning a structure-function analysis on a K^+-specific channel that was first cloned

from a kidney cDNA library and that shows a remarkable molecular property: a very small polypeptide of only 130 amino acids—some 10-fold smaller, for instance, than the voltage-dependent Na^+ channel. We have constructed and expressed a fully synthetic gene for this "mink" channel, using the degeneracy of the genetic code to build a large number of unique restriction sites throughout the coding sequence. Thus positioned to perform routine cassette mutagenesis, we are initiating a search for functional domains of the channel and developing direct tests to settle whether this gene does in fact code for an ion channel at all, a basic question that has not yet been answered rigorously. On the basis of point mutants that lead to changes in ion selectivity, we consider that mink is a structural gene for an ion channel.

High-Level Expression of Ion Channel Proteins

Over the past few years, ion channel genes have been cloned at a rapidly accelerating pace, and they can be expressed in systems that allow high-sensitivity electrophysiological assays. But channels are proteins, and it would be desirable to have a high-level expression system available for producing milligram-scale protein for cloned channel genes. We have found that the baculovirus-Sf9 transient expression system works well with the Shaker K^+ channel. By infecting a cell line derived from the armyworm caterpillar with a Shaker-recombinant baculovirus that normally attacks the alfalfa-looper, we have shown that electrophysiologically normal K^+ currents can be observed and that the Shaker protein is the major membrane protein produced. We are presently testing the wider generality of this system and using this insect-based "channel factory" as a source for purification and functional reconstitution of the Shaker K^+ channel protein.

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 is the Young Investigator Award from the Society for Neuroscience.

MY research concerns the function and development of the visual system, especially the visual areas of the primate cerebral cortex. My 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.

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, 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 defects.

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 and his Ph.D. and M.D. degrees from Harvard University. He was a research fellow at the Whitehead Institute, Massachusetts Institute of Technology, in the laboratory of David Baltimore. He then held a faculty position as an instructor at Harvard Medical School before moving to the University of Michigan.

T lymphocytes protect the body from invasion by foreign organisms, but can also become targets of infection by viruses. One of these is the human immunodeficiency virus (HIV), which causes the acquired immune deficiency syndrome (AIDS). Under normal circumstances, T cells are activated in response to infection and begin to synthesize a set of proteins that step up the immunologic defense system. In T cells that contain HIV, cellular activation signals the virus to augment its replication. We have characterized regulatory proteins that stimulate gene expression in T cells and retroviruses. These cells provide a model for the study of coordinate gene expression during development and viral infection.

Through the use of this model, we have identified proteins that bind to control regions, regulating the expression of other immune system proteins and of HIV. We have also begun to use our knowledge of cellular and viral transcription to deliver recombinant genes *in vivo*. The T cell model has taught us more about the biology of these genes, and we have applied this knowledge to endothelial and vascular smooth muscle cells of the vessel wall, which has provided new opportunities for therapeutic gene transfer.

Regulation of HIV Gene Expression in T Cells and Monocytes

HIV expression can be triggered in T cells by means of phorbol esters or other immune system activators. We have shown that stimulation of these cells increases the binding activity of a protein called NF-κB (nuclear factor that recognizes a sequence in the κ immunoglobulin light chain of B cells), which binds to a DNA control region. This factor stimulates HIV transcription in activated T cells. The DNA sequence that NF-κB recognizes is twice repeated in the HIV control region, and mutation of these sites abolishes HIV inducibility. This transcription factor acts in synergy with HIV products, such as the *tat*-I gene, further enhancing HIV gene expression in an infected cell.

In addition to HIV type 1 (HIV-1), AIDS can be induced by a related virus, HIV type 2 (HIV-2). A retrovirus, HIV-2 shares nucleic acid and protein similarity with HIV-1. First described in West Africa, it has begun to appear throughout the world. The two viruses differ in the length of the asymptomatic period following infection. Because progression of HIV-related disease is associated with increased viral replication, the rate of disease progression may be influenced by virus-activating regulatory proteins synthesized by host cells.

Such proteins could themselves be regulated by distinct cofactors that selectively stimulate cellular activation pathways. These T cell activation pathways regulate specific transcription factors, which may contribute to the regulation of the latent phase of HIV infection.

We have recently defined the transcriptional regulation and induction of these retroviruses and have found that the regulation of HIV-2 differs from that of HIV-1. 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 is recognized by a sequence-specific DNA-binding protein, NF-CD3R.

Jurkat T leukemia cell lines containing HIV-2 provirus also showed increased viral replication following stimulation of the T cell receptor complex, in contrast to HIV-1. These findings suggest that transcriptional regulation and induction of HIV-2 differs from HIV-1. The studies also raise the possibility that different cofactors contribute to activation of disease associated with the two HIV types.

Alteration of gene transcription by inhibition of specific transcriptional regulatory proteins would help to define how these factors participate in cellular differentiation and viral infection. Although several methods are known by which the functions of these proteins can be antagonized, each has specific limitations. Recently we have developed inhibitors of sequence-

specific DNA-binding proteins with double-stranded phosphorothioate oligonucleotides, using octamer or κB consensus sequences. These oligonucleotides bind specifically to the relevant transcription factor. Octamer-dependent activation of a reporter plasmid, or NF-κB-dependent activation of the HIV enhancer, can be appropriately inhibited when the relevant phosphorothioate oligonucleotide is added to a transiently transfected cell line.

Addition of octamer phosphorothioate oligonucleotides to Jurkat T leukemia cells inhibited secretion of interleukin-2 (IL-2) to a degree similar to that observed with a mutated octamer site in the IL-2 enhancer. Double-stranded phosphorothioate oligonucleotides compete for the binding of specific transcription factors and may therefore provide antiviral, immunosuppressive, or other therapeutic effects.

Expression of Cellular and Retroviral-Vector Genes *In Vivo*

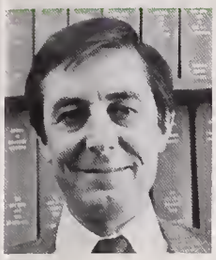
Transcription factors resembling NF-κB also regulate the expression of normal cellular genes. Examination of a growth factor receptor for another immunologic protein, interleukin-2 (IL-2), revealed a site closely related to the regulatory κB sites of HIV. We have shown that this site is recognized not only by NF-κB but by other proteins. We have cloned two genes encoding these proteins and have evaluated their role in the regulation of different cellular genes. We are also studying the activation of another coordinately expressed T cell gene, IL-2, by characterizing proteins that bind to expression-regulating control regions. In particular, we have identified binding proteins that recognize conserved regions by which IL-2 gene expression is negatively regulated.

Despite recent advances in the understanding of eukaryotic gene regulation, the site-specific expression of genes *in vivo* remains a major obstacle to the therapeutic management of human disease. Using our knowledge of retroviral gene expression, we have developed systems utilizing viral vectors to express biologically active proteins in cells and tissues.

For example, we have recently devised a method that allows 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 by direct infection with a retroviral vector or by DNA transfection using liposomes. Several cell types in the vessel wall were transduced with the recombinant gene, including endothelial and vascular smooth muscle cells. Following retroviral infection, a recombinant reporter gene was expressed for at least five months, with no detectable helper virus generated. Recombinant gene expression was limited to the site of infection and was absent from liver, lung, kidney, and spleen.

Thus site-specific gene expression can be achieved by direct gene transfer *in vivo*, and could be applied to the treatment of such human diseases as atherosclerosis, cancer, or AIDS. Biologically active proteins are now being introduced into cells, including 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.

Molecular Analysis of Muscle Contraction



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 is interested in the molecular mechanisms that regulate the production and function of the contractile system in muscle cells. This apparatus is the molecular motor for locomotion and for the heartbeat. Moreover, variations of this contractile system are involved in maintaining the shape and function of nonmuscle cells.

The functional unit of the contractile system is the sarcomere, with its precisely arranged components. Each of these sarcomeric proteins is the product of a small multigene family. Combinations of these protein isoforms can give rise to many different sarcomeres, and this capacity for generating diversity is further expanded by a process of alternative splicing.

Among sarcomeres, significant functional differences are produced through two main mechanisms. One changes the components either at the transcriptional level or by alternative splicing; the other, by regulating the availability of ions in the muscle cells. These two aspects of contractility continue to be the main focus of our research.

Transcriptional Regulation of Contractile Protein Genes

Which sarcomeres are assembled in a given cell at a particular time depends on which contractile protein genes are expressed in the cell. To analyze the mechanisms involved in switching from one gene to another in the same gene family, we have focused on the genes coding for the myosin heavy chain (MHC). Functionally this is the most important component of the sarcomere because it contains the ATPase activity that converts the chemical energy of ATP into mechanical force. We are currently exploring two main questions in the regulation of the MHC genes: What determines that a given MHC gene is expressed in a given cell type at a particular developmental or physiological stage? What determines the level of expression? For these analyses we have concentrated on the two MHC genes expressed in the myocardium, α and β MHC. These genes are particularly advantageous because their expression during development is modulated in response to different hormones and physiological conditions.

We have identified many of the regulatory DNA sequences involved in the tissue-specific expression of these two genes. We have shown that their expression is determined in large part by negative regulatory elements. These elements interact with nuclear proteins that are present in many different cell types but sparse in muscle cells. The transcriptional factors of the MyoD gene family do not seem to play a role in the regulation of these two MHC genes.

In their expression, positive as well as negative factors are essential. The most important positive factor for α MHC is a thyroid-responsive element (TRE). The interaction of this cis-acting element with different forms of receptor for thyroid hormone and retinoic acid has been analyzed in detail. These interactions are required and sufficient to produce most of the gene's known phenotypes. Interestingly, thyroid hormone represses the β MHC gene.

The main positive regulator of β MHC is a muscle-specific enhancer of complex structure, composed of binding sites for muscle-specific and general transcriptional factors. Expression studies of this binding site, together with mobility shift assays, demonstrate that the protein binding to the site is muscle-specific. Using this sequence as a probe to screen a cardiac expression cDNA library, we have cloned a novel transcription factor that belongs to the family of the homeobox genes. Expression of this factor significantly enhances the expression of constructs containing the β MHC enhancer sequence.

The finding of a muscle-specific transcription factor that does not belong to the MyoD family has long been a goal of workers in muscle dynamics. This is so because the MyoD gene family is not expressed in the heart. However, this muscle expresses many of the skeletal muscle genes whose induction is dependent on MyoD. The search for the cardiac MyoD equivalent in many laboratories, including ours, has so far been unsuccessful. The identification and cloning of a cardiac transcription factor of the homeobox fam-

ily suggests that this factor, or a close relative, is a cardiocyte lineage-determining gene. Experiments are in progress to determine its physiological role, its mode of action, and its developmental and physiological regulation.

Regulation of Protein Diversity by Alternative Splicing

To explore the mechanisms involved in the production of different proteins from a single gene, 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. These are produced by the alternative use of two different promoters and poly(A) addition sites, together with three pairs of exons that are used in a mutually exclusive fashion. We have focused on one of these exon pairs (exons 2 and 3) to elucidate the elements involved in this type of regulation.

Using a combination of *in vivo* and cell-free splicing assays, we have analyzed the regulatory features of this system. The mutual exclusive behavior is due to a competition between the two exons for the common flanking splicing sites. Inclusion of exon 3 is the default pattern and occurs in all cell types, except smooth muscle cells. If exon 3 is deleted, however, inclusion of exon 2 becomes the default pattern and occurs in all cell types.

The basis for this behavior is the polypyrimidine tract located at the 3′ end of the intron, between the branch site and the splice site. The role of the tract in splice site selection is determined by its ability to bind to a polypyrimidine binding protein (PBP). This factor has proven to be essential for splicing. During the past year we have biochemically purified and characterized it.

In addition, we have purified and cloned a ribonucleoprotein that copurifies with PBP. Experiments are now in progress to determine the stage in spliceosome assembly that requires PBP as well as its interaction with other components of the splicing complex.

The role of PBP seems to explain the default splicing pattern. However, since PBP is a ubiquitous splicing factor richly present in smooth muscle cells, the obvious question of how the regulated pattern is established remains unanswered. *In vivo* and *in vitro* experiments have clearly demonstrated a negative regulatory mechanism. Factor(s) present in smooth muscle cells block the default splicing pattern, whereupon exon 2 becomes the default pathway and is included in the mature mRNA. The sequences involved in this negative regulation have been mapped to three different elements in and around exon 3. Each of these elements is required, but neither is sufficient to produce this form of regulation. Experiments are in progress to isolate and characterize the trans-acting factors that interact with these elements, using a combination of biochemical and genetic approaches.

Maintenance of the Terminally Differentiated State

One of the more intriguing characteristics of muscle cells is their terminally differentiated phenotype. In the process of differentiation, these cells withdraw irreversibly from the cell cycle and are therefore unable to regenerate. Moreover, expression of muscle-specific genes is dependent on this terminally differentiated state. In its absence, these genes are only transiently expressed and repressed in response to growth factor stimulation. In an attempt to understand this process, we have reversed the terminally differentiated state through the expression of several DNA tumor virus oncogenes. All these oncogenes interact with the product of the retinoblastoma gene. Using antibodies against the transforming protein, Rb, p53, and muscle-specific transcription factors, we have determined the involvement of these several gene products in the production of the terminally differentiated state.

Structure-Function Relationships in Potassium Channels

The contractile cycle of the sarcomere is triggered by the action potential. Potassium channels are fundamental to the repolarization of the cell membrane. For this reason, we recently initiated an in-depth analysis of the structure of a mammalian potassium channel cloned in our laboratory. This channel belongs to the family of delayed rectifiers and has very low inactivation kinetics, making it a valuable model to study the voltage sensor mechanism in this molecule.

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. He completed an internship and residency in medicine at Columbia-Presbyterian Medical Center, New York. 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. He received the Nobel Prize in medicine or physiology in 1978.

THE growth of mammalian cells is regulated by extracellular proteins called growth factors. How these proteins induce cell growth is a key question relevant not only to the regulation of cell proliferation and tumorigenesis but more generally to the response of cells to a variety of extracellular signaling agents, such as developmental factors, classical hormones, cell surface and matrix proteins, and neurotransmitters.

The first step in the stimulation of cell growth by a growth factor is the interaction of the factor with a specific cell-surface receptor. This interaction rapidly induces a cascade of biochemical reactions in the cell, leading to the sequential activation of specific cellular genes and subsequent DNA replication and cell division. The ordered expression of cellular genes induced by growth factors has the attributes of a regulated genetic program. Research in my laboratory concerns the analysis of this program and its role in cell proliferation.

To identify some of the genes activated by growth factors we used recombinant DNA methods to isolate DNA copies (cDNAs) of gene messages present in cultured fibroblastic mouse cells only after they have been stimulated by a growth factor or by serum that is rich in a growth factor derived from blood platelets. We and others have identified one set of genes, the immediate-early genes, which are activated within 2 or 3 minutes after addition of growth factor, coordinately with activation of the proto-oncogenes *fos* or *myc* detected previously. Some of these immediate-early genes encode proteins related to known transcription factors (proteins that regulate gene expression), others encode secreted proteins or membrane proteins, and others encode proteins that are part of the filamentous structures of cells. We have concentrated largely on genes that encode probable transcription factors, because these are likely to be involved in regulating the genetic program induced by growth factors.

We previously described immediate-early transcription factors with "zinc finger" motifs and others that are "leucine zipper" proteins of the Jun family. During the past year we have identified three additional immediate-early proteins whose structures are related to those of known transcription factors. One of the proteins is Nup475, a nuclear, zinc-binding protein that has two copies of a sequence that suggests it can form a novel type of zinc finger. Studies are under way to determine if Nup475 binds to a specific DNA sequence and regulates transcription.

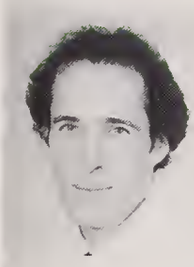
The second protein, ΔFosB, is a naturally occurring shortened form of FosB, a previously described member of the Fos proto-oncogene family of transcription factors. Expression of ΔFosB was found not only in growth factor–stimulated cultured cells but also in specific parts of the brain after electrical stimulation and in regenerating liver. Like other members of the Fos family, ΔFosB can form dimers with members of the Jun family, and the dimers can bind to signals in DNA. However, unlike other Fos-Jun dimers, ΔFosB-Jun dimers are not transcriptionally active. Instead, ΔFosB competitively inhibits the transcriptional activities of Jun and Fos; it may therefore serve as a negative regulator of Jun and Fos during the growth response.

The third protein is HLH462, which is related to the helix-loop-helix class of transcription regulators. Proteins of this class form dimers through their helix-loop-helix structural domains. Some dimers bind to signals in DNA, whereas others cannot bind to DNA because they lack a DNA-binding region. Helix-loop-helix proteins of the latter type have been shown to inhibit the transcriptional activity of DNA-binding helix-loop-helix proteins. HLH462 has the structural and functional properties of an inhibitory helix-loop-helix protein. It is expressed in many mouse tissues and in early mouse embryos. What role it plays in the growth factor–induced program is not yet clear. One possibility is that it inhibits a negative growth regulator present in nongrowing cells.

So far, approximately 15 known or probable transcription factors have been identified among the immediate-early proteins induced in fibroblasts by serum growth factors. Later in the cellular response other genes are activated whose encoded proteins are thought to mediate progres-

sion toward the DNA synthesis phase of cell growth. The diversity of immediate-early transcription factors probably reflects a complex mechanism for activating these later genes.

Molecular Biology of Visual Pigments



Jeremy Nathans, M.D., Ph.D.—Assistant Investigator

Dr. Nathans is also Assistant 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. 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. It 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 taking three general approaches related to the visual pigments: investigations of their structure and function, the control of their expression, and their variation within the human population.

Structure/Function Studies

Several years ago we succeeded in producing large quantities of bovine rhodopsin by expression of cloned cDNA in tissue culture cells. We are using this system in conjunction with site-directed mutagenesis to define the chromophore binding pocket, the residues involved in protein conformational changes, and the surface and transmembrane topography of the protein.

In one experiment, each of the 22 negatively charged amino acids (e.g., aspartate or glutamate) was changed to a neutral residue of identical size (asparagine or glutamine). We observed that only one of the mutant proteins differs dramatically from the wild type in its behavior. When glutamate^{113} is mutated to glutamine, the Schiff's base linking 11-cis retinal to the protein loses its proton. The base is re-protonated upon addition of small anions (e.g., chloride) to the sample.

Our interpretation of this experiment is that glutamate^{113} is normally the counterion that stabilizes the protonated Schiff's base, but in its absence a small anion can serve as a surrogate. This observation is likely to be relevant to the mechanism of protein activation, because Schiff's base deprotonation is required for the protein to assume its active conformation.

In a second experiment, Charles Weitz, a postdoctoral fellow, has examined a number of rhodopsin mutants for their ability to assume the active conformation. Thus far, one mutant binds the chromophore and absorbs light normally but appears to be locked in the inactive conformation.

In a third set of experiments, Jimo Borjigin, a graduate student, is using insertional mutagenesis to tag rhodopsin at predetermined sites. The tags consist of stretches of foreign amino acids that can bind to other proteins (e.g., antibodies). These modified rhodopsins should be useful for a variety of structural studies.

The expression system is also being used by Shannath Merbs, a graduate student, to produce the human cone pigments—a related group of light receptors that mediate color vision.

Control of Visual Pigment Gene Expression

As an entree into the general question of retinal development, we are examining the control of the genes for visual pigment. Donald Zack, a postdoctoral fellow, in collaboration with Jean Bennett and John Gearhart at Johns Hopkins, has constructed a set of transgenic mice that carry sequences upstream of the bovine rhodopsin gene joined to a gene encoding β-galactosidase, a convenient histochemical marker. Rhodopsin gene fragments as small as 230 base pairs direct expression of the reporter gene to the photoreceptor cells.

Curiously, a considerably larger fragment (2,100 bp) directs expression in a spatially nonuniform manner: a gradient of expression forms across the retina. This DNA fragment appears to be responding to a preexisting spatial gradient. This gradient may be involved in determining the

map coordinates for the wiring of the visual system.

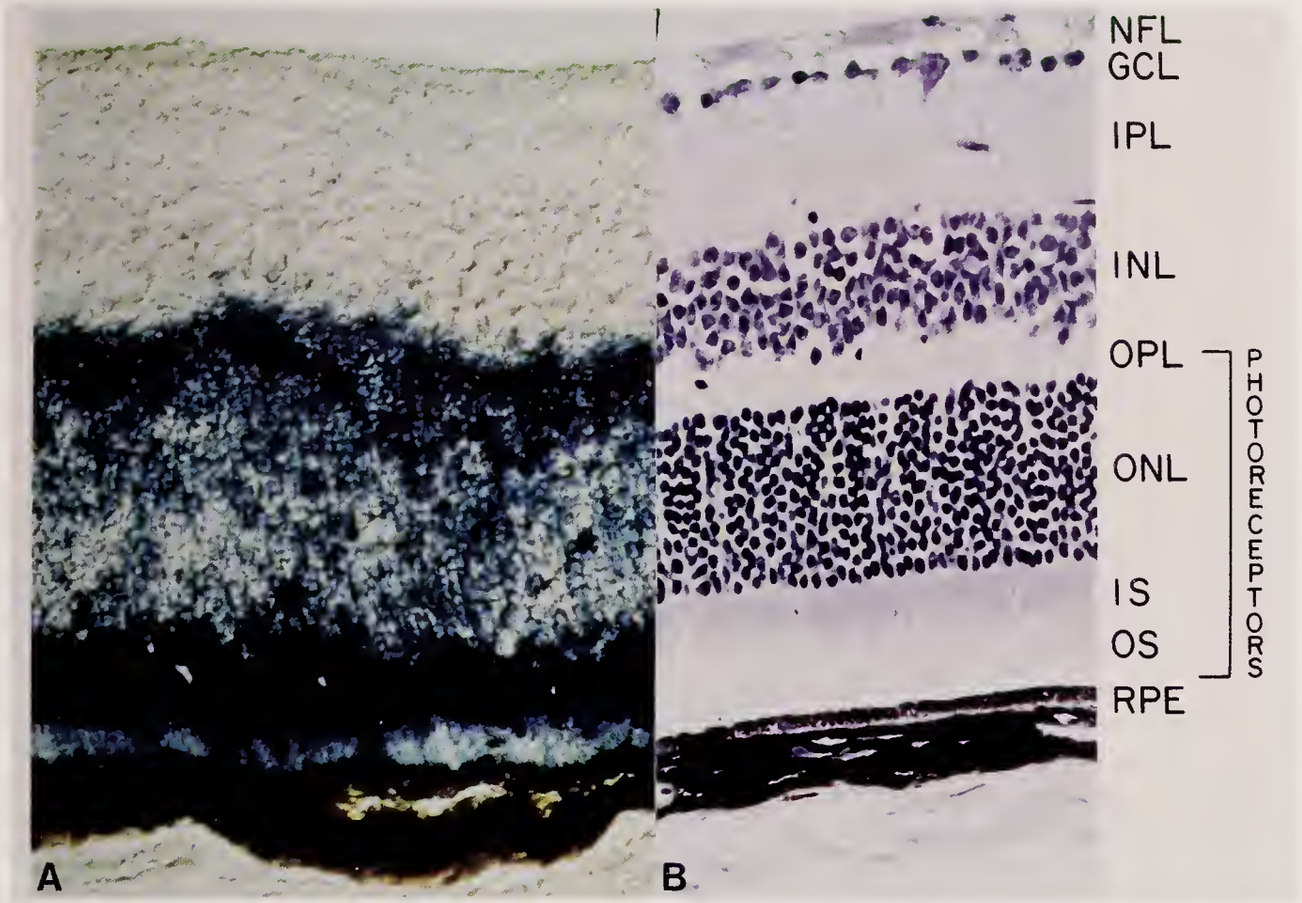
Inherited Variation

The human visual system has several experimental attributes that facilitate a genetic analysis: first, humans can accurately report their visual experience in sensitive, noninvasive behavioral tests; second, inherited alterations in visual function rarely affect longevity or fecundity—and thus usually persist in the gene pool; and third, persons with significant visual disorders usually present themselves to the medical community. We have taken advantage of these attributes to identify persons with mutations in each of the four genes for visual pigments.

Several years ago we showed that the common forms of red-green colorblindness are due to mutations in the red and green cone pigments. More recently Charles Weitz has identified amino acid substitutions in the blue pigment gene in patients with defects in blue sensitivity. A rare form of severe color vision deficiency, blue cone monochromacy, occurs when both red and green cone systems are inactive. Our analysis of 30 blue cone monochromat families shows that deletion of a small DNA sequence adjacent to the red and green pigment genes suffices to cause the defect. Yanshu Wang, a graduate student, has recently shown that a large DNA fragment containing this region directs expression of a reporter gene to cone photoreceptor cells in transgenic mice.

To identify rhodopsin gene mutations, we examined patients with defects in night vision and peripheral vision. (Rhodopsin is the visual pigment in rods—the photoreceptors, enriched in the peripheral retina, that subserve dim-light vision.) Night blindness and loss of peripheral vision are the hallmarks of retinitis pigmentosa, a family of genetic disorders that affect 1 person in 4,000. We recruited patients with retinitis pigmentosa, and Ching-Hwa Sung, a postdoctoral fellow, screened their rhodopsin genes, using denaturing gradient gel electrophoresis, a new and rapid method for detecting point mutations.

Thus far we have discovered 13 different point mutations in the rhodopsin gene, together accounting for 24 percent of patients with autosomal dominant retinitis pigmentosa. All of the mutant proteins have been produced in the tissue culture expression system. The single most common mutation, a proline-to-histidine substitution at amino acid 23, results in a protein that becomes stuck in the endoplasmic reticulum.



Sections of a transgenic mouse retina expressing β-galactosidase under control of a bovine rhodopsin promoter. A: The β-galactosidase substrate X-gal stains only the photoreceptors. B: Hematoxylin and eosin staining shows the cell layers.

From Zack, D.J., Bennett, J., Wang, Y., Davenport, C., Klaunberg, B., Gearhart, J., and Nathans, J. 1991. Neuron *6:187–199.*

Gene Regulation in Animal Cells

Joseph R. Nevins, Ph.D.—Investigator

Dr. Nevins is also Professor of Microbiology and Immunology and of Genetics 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.

THE regulation of gene expression is central to the complex cellular changes that take place during such events as embryogenesis and oncogenesis. An elucidation of the molecular mechanisms of gene control pathways is the focus of our laboratory, including the modification and regulation of factors that mediate the control of gene expression.

Mechanisms of Viral-mediated Trans-activation of Transcription

The primary event in the expression of a gene is the initiation of transcription leading to the formation of functional mRNA. The control of this event is a key step in the determination of cellular phenotype. Various studies have shown that transcription initiation is complex, resulting from the interaction of a large number of proteins with regulatory DNA sequences of the gene.

The study of complex cellular events is often facilitated by the use of simple viral model systems. An example is the trans-activation of transcription by viral regulatory proteins such as the adenovirus E1A gene product. The E1A protein mediates the activation of transcription of a set of viral genes and several cellular chromosomal genes. Since the activation process involves the use of cellular transcription factors, this system has provided a means for the study of cellular transcription control.

Work in our laboratory has led to the identification of several cellular transcription factors that the viral regulatory system utilizes and that contribute to the trans-activation event. In particular, the activity of these factors is modified by the viral infection, which increases the levels and changes the nature of the DNA-binding activity of the proteins. The DNA-binding activity of several factors depends on the phosphorylation state of the protein, and phosphorylation appears to play a key role in the activation process.

Recent studies have shown that the E2F transcription factor, as isolated from adenovirus-infected cells, binds to the E2 promoter with a high degree of cooperativity, resulting in the formation of a very stable DNA-protein complex. This binding is important for full transcription activation, requires the precise dyad arrangement of the E2F sites as found in the E2 promoter, and requires the interaction of a 19-kDa product of the early viral E4 gene with the E2F factor. The induction of cooperative E2F binding is striking in view of the specificity. Although E2F-binding sites have been identified in a variety of cellular promoters, no other instances have been found of this precise arrangement in the E2 promoter. It thus appears that the cellular E2F factor is converted into an E2 promoter-specific factor through the action of the E4 protein.

The E2F factor has now been found to be complexed to cellular factors in extracts of a variety of cell lines. These complexes are significant with respect to a viral infection, since the E4 protein cannot interact with E2F that is already complexed to a cellular factor, thus preventing the formation of the stable interaction on the E2 promoter. Strikingly, however, we have found that E1A proteins can dissociate these E2F-containing complexes, releasing free E2F that can associate with the E4 protein. Thus it would appear that a two-step mechanism has evolved to utilize the cellular E2F factor efficiently and redirect it for viral-specific purposes.

Through the analysis of a large series of E1A mutants it has become clear that the E2F-dissociating activity is dependent on sequences within E1A that are important for transforming activity. Moreover, trans-activation assays have demonstrated that mutations that disrupt E2F dissociation also disrupt cell-specific trans-activation function. These are also the amino acid sequences that are shared with two other viral oncogene products—SV40 T antigen and human papillomavirus E7. Previous experiments have shown that both possess trans-activation function dependent on the E2F factor. Recent experiments have shown that the E7 protein is capable of altering the E2F complexes in a manner similar to E1A. It therefore appears that these otherwise unrelated viruses have acquired a common regulatory function through evolution, directed at a specific cellular transcription factor, that may be

a critical part of the ability of these viral proteins to induce oncogenic transformation.

Gene Regulation by Alternative RNA Processing

Very often a transcription unit encodes not one mRNA but several (and thus several proteins). By alternative processing of the primary transcript, one particular mRNA is selected and produced. In several instances, this is a regulated event, subject to change depending on the circumstances of the cell. Alternative processing can involve both the selective splicing of exons and the selection of one of several poly(A) addition sites, thus generating a unique mRNA 3′ terminus.

Possibly the best example of regulation of poly(A) site utilization is within the differentiating B lymphocyte, where there is a dramatic change in the nature of the immunoglobulin (Ig) heavy-chain polypeptide that is synthesized. Initially, a membrane-bound form of the protein is produced, but this then switches to a secreted form when the cells differentiate to plasma cells. The RNAs that direct the synthesis of these two proteins are encoded in the same transcription unit and are produced by differential cleavage of the primary transcript at two different poly(A) sites.

A definition of the events and factors involved in poly(A) site formation, again using simple viral systems but also the developmentally regulated Ig heavy-chain gene system, has been a major goal of our laboratory. Experiments have identified sequences within the two Ig μ heavy-chain poly(A) sites that are essential for regulated selection during B cell differentiation. In addition, we have utilized cell-free systems capable of accurate poly(A) site processing *in vitro* to identify and isolate factors mediating the processing reaction.

Purification of factors from these cell-free extracts has now yielded information concerning the multiplicity and nature of factors involved in this processing event. At least four distinct activities are required for efficient processing at the poly(A) site. Two of these factors form complexes with the pre-mRNA, and the combined interaction results in the generation of a stable, committed complex that allows processing to take place. Furthermore, the stability of these complexes can vary, depending on sequences in the RNA that are known to be important for efficient processing, indicating that the formation of these protein-RNA interactions plays a significant role in determining the efficiency with which a particular poly(A) site is utilized.

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 generally unknown. Recombinant DNA techniques are being employed to isolate the responsible genes, with the aim of furthering our understanding of the normal processes that result in each of these diseases when disrupted.

Fragile X Syndrome

One of the most common genetic forms of mental retardation, the fragile X syndrome, affects nearly 1 in every 2,000 males worldwide. The chromosomes of affected males are normal in appearance on routine examination. If, however, supplies of the molecular building blocks for DNA synthesis are stringently limited during cell division (*in vitro*), an abnormal gap or fragile site is induced to appear at the tip of the human X chromosome.

The disease is generally inherited in an X-linked manner, but there are important exceptions. These are found in "transmitting males," the approximately 20 percent of male carriers who are not retarded and do not show the chromosomal fragile site under the usual induction conditions. The grandsons of transmitting males, however, may inherit the gene through daughters and manifest the disease.

Understanding the fragile X syndrome will ultimately require isolation of DNA sequences from the region involved so that the molecular basis of both the chromosomal abnormality (the fragile site) and the clinical disease (fragile X mental retardation) can be elucidated. My colleagues and I are creating yeast artificial chromosomes (YACs) containing DNA from the region of the fragile site for study. This involves attaching functional parts of yeast chromosomes to large fragments of DNA from a patient and reintroducing these fragments into yeast. The large fragments are then propagated in the yeast, just like a yeast chromosome, and can be grown in large amounts for study. Nearly 600,000 bp of contiguous DNA from the region around the fragile site has been isolated to date. Our aim is to understand how the chromosome abnormality comes about and what gene or genes are responsible for the retardation.

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 been studying a female patient in whom choroideremia has occurred because of a disruption of the choroideremia gene caused by a chromosome break in the X chromosome in this region. A transcribed gene that is disrupted by this chromosome 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 gene identified in both laboratories bears no sequence homology to any previously identified gene, and there is no information at present as to what the protein encoded by this gene does and why mutations in the gene cause choroideremia.

The goal is to learn more about the function of the normal retina as well as to find new methods of diagnosis and treatment for this and other related retinal disorders.

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.

Linkage analysis with restriction fragment length polymorphisms localized the gene to the Xq25 region of the X chromosome. Collaborators in the United Kingdom have identified a Lowe's syndrome patient in whom the disease is caused by a break in Xq25 that has disrupted the gene, splitting it into two pieces separated on different chromosomes. In collaboration with David Nelson (Baylor College of Medicine), we have identified a segment of DNA in a YAC that spans the region of the break in this patient. Segments of DNA from within this YAC have been shown to contain gene sequences that are candidates for containing the Lowe's syndrome gene.

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.

Function of Proto-oncogenes in Early Embryogenesis

Roeland 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. Last year he moved to the Beckman Center of Stanford University and the Howard Hughes Medical Institute. He is a member of the European Molecular Biology Organization (EMBO).

THE major goal of our work is to elucidate how intercellular signals control the proper arrangement of cells and tissues during early embryogenesis. These signals can be in the form of polypeptides that are secreted from one cell and received by others through binding to specific receptors. A systematic approach to the identification of communication molecules involved in embryogenesis is only possible in organisms where saturation mutagenesis can be applied to screen for genes with a relevant phenotype. Both in *Drosophila* and in *Caenorhabditis elegans*, such approaches have led to the discovery of a series of developmental genes, whose functions can be analyzed by molecular cloning. In combination with cell transplantation and cell ablation experiments, this method is extremely powerful and has led to illuminating insights.

From another line of research, it has become clear that some proteins originally identified as growth factors in adult organisms also act as important regulators early in embryogenesis. And other molecules implicated in embryonic signaling have turned out to be the products of proto-oncogenes: genes that normally have essential functions in the regulation of cell proliferation, but whose altered expression can lead to cancer.

The *Wnt/wingless* gene family is one of the best examples of the link between cancerous growth and the control of normal development. The prototypic member of this group is *Wnt*-1, an oncogene frequently activated in mouse mammary cancer. The *Wnt*-1 gene is normally not expressed in mammary gland or in most other adult tissues, but its transcription can be induced by nearby insertion of proviral DNA of a retrovirus, the mouse mammary tumor virus. *Wnt*-1 encodes a secreted protein rich in cysteine residues, and the gene is normally expressed only during early mouse embryogenesis, in particular in the closing neural tube and the developing brain. Proof that *Wnt*-1 is an oncogene came from transfection experiments and from the finding that, placed as a transgene in the germline of mice, it can lead to tumor induction.

Recently we and others have found that *Wnt*-1 is part of a gene family comprising, in the mouse, at least nine members. Most of these *Wnt*-1-related genes have a very restricted pattern of expression during early developmental stages, and at least several members of this group can behave as oncogenes when activated by insertion of proviral DNA in mouse mammary tumors. The *Wnt*-3 gene, for example, is more than 50 percent identical to *Wnt*-1 and is activated in a low percentage of mammary tumors.

We have performed a 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 some areas that overlap the expression domain of *Wnt*-1 but also in unique domains. 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. This could now be tested by generating mice with mutations in these genes, using homologous recombination in embryonic stem cells.

Our present aim is to understand the mechanism of action of the *Wnt*-1 gene family during embryogenesis and to extrapolate these findings to cancerous growth. We wish, for example, to identify the receptors for these secreted molecules. One of our model systems is the fruit fly *Drosophila*. Some years ago we made the surprising observation that the homologue of the *Drosophila Wnt*-1 gene was identical to the segment polarity gene *wingless*. Because of the extensive genetic analysis of *Drosophila* embryogenesis, in particular the mechanism of segmentation, this observation has allowed us to study the interactions of *Wnt*-1/*wingless* with other genes.

The basic body plan of the fruit fly is set up 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. The gap genes and the pair-rule genes encode nuclear proteins and are active before the *Drosophila* embryo becomes cellularized, and the segment polarity genes are the first ones 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 the protein in whole-mount embryos and in individual cells.

The protein is seen on the surface of cells and in intracellular structures that constitute uptake vesicles. Such structures are also seen in cells adjacent to those that make the *wingless* protein, suggesting a paracrine mechanism of action of the gene. When we now look at the distribution of the *wingless* protein in embryos that are mutant for some other segment polarity genes, we can observe differences suggesting that *wingless* interacts directly with the products of these genes.

In other experiments, we have overexpressed the *wingless* gene from a heat-shock promoter and transfected it into established *Drosophila* cell lines. By deliberately altering the expression of the gene in whole embryos or in cultured cells, we are identifying additional genes that are involved in the *wingless* signal transduction pathway. For example, our finding that misexpression of *wingless* in an embryo results in large areas of pattern abnormalities can be taken as evidence that many, if not all, cells can receive the *wingless* signal and must therefore have receptors.

We have also found that *wingless* in *Drosophila* is also 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, but in characteristic patterns that differ from *wingless*.

In the analysis of the mechanism of action of the *Wnt* genes in mouse development, we hope to take advantage of the results of the *Drosophila* work, conceivably by isolating the mouse homologues of those fly genes that have been shown to interact with *wingless*.

Molecular Genetics of B Cell Development

Michel C. Nussenzweig, M.D., Ph.D.—Assistant Investigator

Dr. Nussenzweig is also Assistant Professor at the Rockefeller University. He received his undergraduate and his 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 the Harvard Medical School with Philip Leder.

THIS laboratory is developing ways of understanding the cells of the immune system at the molecular level, using the tools of molecular biology and genetics.

The immune system is responsible for protecting vertebrates from both invasion by infectious organisms and deregulated growth of endogenous malignancies. In order to accomplish this task, the 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 diversity in their surface receptors 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, the heavy- and light-chain immunoglobulin proteins that serve as the B cell receptor derive from two sets of rearranging genes. The same genes also direct the production of secreted antibodies that are an important component of the humoral immune system. 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 the 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. In order 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 results in the exclusion of most endogenous mouse immunoglobulins. The secreted version of the same transgene had little effect.

In order to examine the molecular basis for exclusion by membrane-associated IgM, we have developed a sensitive PCR assay for heavy-chain gene recombination. This assay allows us for the first time to examine specific heavy-chain recombination events in pools of primary lymphoid cells. We are able to assess D to J segment recombination as well as V to DJ joining events for specific families of variable regions.

We find that the rearrangement of the endogenous heavy-chain locus is inhibited in transgenic mice that carry human membrane IgM. Gene rearrangements are not affected in the control transgenic mice that carry the secreted human IgM gene. Furthermore, not all gene segments are equally affected by the membrane transgene. The large families of variable-region genes that are in the distal part of the immunoglobulin locus are almost completely inhibited from undergoing rearrangements, whereas the small families of more

proximal variable-region segments, as well as the D and the J segments, are not affected. We are currently investigating the possibility that this positional effect is related to transcription and proximity to the immunoglobulin heavy-chain enhancer.

In another area of investigation suggested by our initial experiments, we are trying to understand how a membrane-anchored immunoglobulin produces a signal. Although the structure of this receptor has been known for several years, the mechanism of signal transduction has not been elucidated. Recent work has implicated tyrosine phosphorylation as well as GTP hydrolysis in the signaling event, and a series of immunoglobulin-associated proteins have been identified. We have started to pursue this problem with the goal of developing a cell-free system that will allow us to isolate and reconstitute the functional components.

Mechanism of DNA Replication

Michael E. O'Donnell, Ph.D.—Assistant Investigator

Dr. O'Donnell is also Assistant Professor of Microbiology at Cornell University Medical College, New York City. He received his Ph.D. degree with Charles H. Williams, Jr., from the Department of Biological Chemistry at the University of Michigan, Ann Arbor, on electron transfer in the flavoprotein thioredoxin reductase. 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, 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.

The aim of our research 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 *E. coli* chromosome is replicated by over a dozen proteins. Ten of these are tightly bound into a complex that contains the DNA polymerase activity. Our present research is focused on this complex, called DNA polymerase III holoenzyme, which will be referred to below as "the holoenzyme."

Only two proteins of the holoenzyme have well-defined functions: α, the DNA polymerase protein, and ϵ, an exonuclease that proofreads the product of the polymerase protein. Our aim is to determine the individual functions of the other eight "accessory proteins" of the holoenzyme. We hope that analysis of the *E. coli* holoenzyme will extend and generalize the understanding of the replication process in all organisms.

We have recently developed methods to obtain nearly pure preparations of each protein, or subunit, of the *E. coli* holoenzyme, 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, hydrolyses 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 DNAs.

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). Hence three accessory proteins of the holoenzyme (γ, δ, β) are needed to confer rapid and processive synthesis onto the polymerase subunit. This fits nicely with the fact that the *E. coli* cell duplicates its 4 million–base chromosome within 30 minutes.

After the holoenzyme has replicated the DNA, it remains bound to it. However, upon addition of primed DNA containing a β clamp, the new clamp specifically seeks out polymerase molecules bound to completed DNA templates, and the β clamp on the fresh primed DNA "steals" the polymerase away from the clamp on the completed template. Hence the β protein appears to accelerate the polymerase as well as mediate its rapid cycling from a completed DNA to a new one. The rapid cycling 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 be used over and over (cycling) every 1–2 seconds.

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 four (γ, δ, β, τ) of the eight accessory proteins of the holoenzyme 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.

Our future plans include assignment of function to the other four accessory proteins, and identification of the genes for the remaining five accessory proteins of the holoenzyme in order to obtain them in large quantity. Possession of the proteins in quantity would allow us to start investigations on the structure of the holoenzyme. Identification of their genes would also allow construction of *E. coli* genetic mutants, which would provide further clues to the exact function of each subunit in replication. Clearly, many more studies by our laboratory and others are needed to gain a full understanding of the mechanism by which duplex DNA is replicated.

Dealing with DNA on a Large Scale

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 (bp) 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 remains 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.

The heart of our approach is a new method of cloning large DNA molecules from any organism. DNA cloning, which was the root technology of the recombinant-DNA revolution of the 1970s, depends on splicing *vector* sequences onto other DNA molecules to create a new replicon—that is, a DNA molecule that will replicate inside a host cell, allowing large numbers of identical copies to be produced in cell culture. The replicons in the new cloning system are yeast artificial chromosomes (YACs). DNA sequences in the YAC vectors impart the properties of a true yeast chromosome to the foreign DNAs to which they are spliced. Once introduced into host yeast cells, the YACs replicate at each cell cycle during the growth of the host and segregate faithfully into the two progeny cells.

One advantage of YACs over previous cloning systems is that there is no absolute upper size limit. At present, we can 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 cloning systems. Another advantage is that the methods of packaging, maintaining, and replicating DNA are more similar in yeast than in bacteria to the analogous methods in cells of higher organisms. Consequently, it will likely be possible to propagate more segments of human DNA in yeast cells than in bacterial hosts.

The first phase of this project has involved establishing basic "library" technology for the YAC system. We have prepared more than 80,000 YAC clones containing segments of human DNA that average 250,000 bp in size. A typical segment of human DNA is represented six to seven times in this library on YACs of independent origin. We have also developed efficient screening methods, based on the polymerase chain reaction (PCR), which allow YACs containing particular segments of the human genome to be identified in the library. Screening can be carried out even if one's only prior knowledge of the segment is the sequence of a mere 100–200 bp of the DNA.

The YAC library has yielded clones containing more than 200 different human genes. These clones are being employed in many collaborating laboratories for such purposes as aiding in the search for genes involved in specific human diseases. Examples include the successful search during the past year for the gene that is mutated in neurofibromatosis, carried out in part in the laboratory of Francis Collins (HHMI, University of Michigan). Andrew Feinberg (HHMI, University of Michigan) has also used YACs to characterize a large region of chromosome 11 that is implicated in the etiology of a childhood cancer, Wilms' tumor, as well as several developmental defects. YACs are also playing a key role in the worldwide effort to identify the gene that is mutated in Huntington's chorea, a fatal, adult-onset neurological disease.

Current research is focused on improving the power of the YAC system for the functional and structural analysis of human DNA. Functional studies depend on the transfer of YAC clones

from yeast into human cells. We have concentrated on YACs containing two human genes, HPRT and GART. Both code for enzymes required for the synthesis of nucleic acid precursors in human cells. Rodent cell lines are available that do not produce these enzymes. We have succeeded in transferring YACs containing the human HPRT and GART genes into these rodent cells and have shown that the genes are expressed. A variety of gene transfer methods have been used, including microinjection and direct fusion of yeast spheroplasts (cells from which the walls have been removed) with rodent cells.

Physical analysis of DNA cloned into YACs also remains a major research interest. YACs have greatly improved our ability to recover DNA from higher organisms in megabase-pair blocks. However, much of the biological interest in these blocks occurs at the level of the base pair, where a single change can have profound consequences. The problem of analyzing megabase-pair blocks of DNA at base-pair resolution remains formidable, particularly since there is a need for generic solutions that can be applied to any large segment of DNA from any organism. Furthermore, given the amount of DNA present in the genome of higher organisms and the central role that its analysis plays in biomedical research, it is essential that these techniques rely primarily on advanced instrumentation rather than skilled laboratory personnel. The challenges and opportunities in this area are reminiscent of those faced by digital computing in the 1950s: promising technology was already in hand, but vastly improved efficiency was required before its promise could be broadly realized.

Molecular Genetics Studies on 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 the 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 Past President of the American Society of Clinical Investigation. Dr. Orkin was recently elected to the National Academy of Sciences.

ALL blood cells derive from pluripotent stem cells in the bone marrow. The decision of stem cells to differentiate leads to the production of a heterogeneous population of cells with varying developmental potentials and commitment to expression of lineage-specific protein products. A goal of this laboratory is an improved understanding of hematopoietic development and the expression and function of specific genes that relate to the normal biology of hematopoietic cells.

Our efforts are concentrated on the analysis of both red and white blood cells. These cell types are important in severe, clinically significant human genetic disorders in which the capacity to produce specific proteins is impaired by mutation. We seek to understand the molecular basis of these inherited disorders, delineate the normal regulation of affected genes, and utilize what is learned to formulate novel treatments based on molecular biologic considerations.

One of the major disorders of red blood cells is β-thalassemia (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. The major, unsolved problems now relate to how globin genes are normally regulated in the developing erythroid precursor cells. Specifically, how are globin genes activated only in red cells but not in other tissues? How are different globin genes regulated in development? To examine these general issues 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 discovered that recognizes a small DNA motif found in the promoters or enhancers of virtually all erythroid-expressed genes, and the human, mouse, and frog homologues were cloned. The protein is modular, consisting of a novel two-finger structure required for DNA binding and other domains that serve as potent activators of gene transcription. The expression of this 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.

An understanding of how this transcription factor is itself regulated in erythroid cells may provide important insights into the initial events in erythroid decision making and maturation. Recent findings indicate that the gene is activated early in hematopoietic development and is later subject to positive autoregulation of its promoter. Site-specific gene disruption in mouse embryo stem cells and generation of chimeric animals has also revealed that the protein is essential for normal erythroid differentiation and that other proteins binding to the GATA motif cannot compensate for its absence. The focus of our studies is to understand in detail how this transcription factor functions in normal erythroid development and how its expression is first turned on in early progenitor differentiation. Ultimately these studies may provide new clues to differential regulation of globin genes and the prospects for directed manipulation of their expression for the treatment of hemoglobinopathies.

In a separate but conceptually related group of studies a gene that encodes an essential component of the white blood cell (phagocytic) system responsible for killing ingested microorganisms is being examined. We wish 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 chronic granulomatous disease, an X-linked condition. By reverse genetics (positional cloning) we previously isolated the relevant gene, determined its structure, and demonstrated the presence of protein product in the cytochrome complex of phagocytic cells. In addition, because interferon-γ stimu-

lates phagocytic cell function generally as well as expression of the cytochrome, it was also possible to show that this lymphoid is clinically effective in chronic granulomatous disease. Studies have identified several point mutations in the cytochrome that interfere with protein function *in vivo*. We are now working to define the elements of the gene responsible for cell-restricted (i.e., phagocytic cell) gene expression. Through the use of transgenic mice, we identified a DNA fragment sufficient for targeting reporter or oncogenes in phagocytic cells. Coupled to an oncogene, this fragment leads to the development of an inherited malignancy of phagocytes in mice.

Efforts now are focused on the characterization of the cellular components that account for lineage-specific regulation. In the end, through an understanding of normal regulation in white blood cells, we envision improved capabilities to modulate gene expression and differentiation in both health and disease. This will facilitate ongoing strategies to develop new approaches to the treatment of genetic disease by gene transfer into marrow stem cells (somatic gene therapy).

Eye Development, Pigmentation, and Insertional Mutagenesis

Paul A. Overbeek, Ph.D.—Assistant Investigator

Dr. Overbeek is also Assistant Professor in the Department of Cell Biology, Institute for Molecular Genetics, and Division of Neuroscience 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 the NIH.

TRANSGENIC mice are used in my laboratory as a model system to study mammalian gene regulation and embryonic development. Such mice are generated by microinjection of purified fragments of DNA into mouse embryos at the one-cell stage. The embryos are then transferred into the reproductive tracts of suitably prepared female mice and allowed to develop to term. In a typical experiment, 20–30 percent of the newborn mice are found to have integrated the injected DNA stably into their genome. The new DNA is termed transgenic DNA.

Through recombinant DNA technology, the DNA for microinjection can be specifically designed and assembled from previously characterized fragments of DNA. The microinjected DNA, once it has integrated, is duplicated along with the rest of the genome at each cell division and is therefore present in all cells of the transgenic mice. The mice are analyzed in detail to determine whether the new DNA causes changes in the development or behavior of the mice.

The new DNA can work through at least two different mechanisms. If the new DNA is expressed—i.e., if it encodes a protein—the characteristics of the mice can be altered in a "dominant" fashion. Alternatively, the transgenic DNA may cause a "recessive" mutation as a result of its novel position in the genome. When the new DNA integrates, the normal organization of the adjacent endogenous genes may be disrupted, resulting in recessive inactivation of those genes. Transgenic mice are typically analyzed for both dominant and recessive changes in their pattern of development.

Much of our research is directed at studies of eye development. The lens of the eye is known to express a set of proteins called crystallins. The crystallins are essential for normal transparency of the lens, and certain crystallins are expressed exclusively in the lens, implying that lens-specific regulatory sequences must be associated with those genes. In collaboration with Joram Piatigorsky and Heiner Westphal at the National Institutes of Health, experiments were done to identify lens-specific regulatory sequences. Recombinant DNA techniques were used to link a putative crystallin regulatory sequence to coding sequences from a bacterial gene. The recombinant DNA construct was used to generate transgenic mice. When the mice were assayed for expression of the bacterial gene, activity was detected only in the lens. This result demonstrated that a small stretch of DNA located near the transcription initiation site of a crystallin gene was sufficient to provide lens-specific regulation of gene expression.

A subsequent experiment was done to generate transgenic mice that would express an oncogene in their lenses. An oncogene encodes a protein that is thought to be capable of predisposing cells to cancer formation. A construct was made that contained the crystallin regulatory sequences linked to an oncogene from a tumor virus, and transgenic mice were generated. The mice displayed dramatic bilateral cataracts, and their lenses were found to contain rapidly growing tumor cells. When the embryonic development of the lenses was studied, it became apparent that tumor cell proliferation began shortly after the onset of expression of the oncogene. These experiments showed that expression of the oncogene was sufficient to induce tumor formation. These transgenic mice provide a model system in which to study the mechanisms by which an oncogene can alter cellular differentiation and proliferation.

Over the past two years, regulatory sequences that give cell-specific gene expression in other regions of the eye have been characterized. In collaboration with Gregory Liou, a promoter that is active in the photoreceptor cells of the retina has been identified. The promoter comes from a gene that encodes a retinoid-binding protein. This protein plays a role in the recycling of retinal, a vitamin A derivative that is essential for normal vision. The identification of a photoreceptor-specific promoter will allow future studies of retinal function and development.

Studies have also been done to identify a promoter that is active in the pigmented regions of the eye. In both mice and humans, loss of pig-

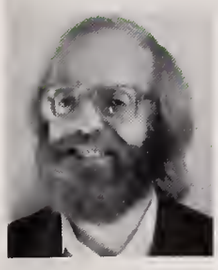
ment synthesis results in albinism. Albinism is often accompanied by a loss of activity of the enzyme tyrosinase, suggesting that albinism might be due to a mutation in the tyrosinase gene. To test this hypothesis, we constructed a nonmutant tyrosinase minigene and introduced it into the genome of albino mice. The resultant transgenic mice were pigmented and could be identified at birth by simple visual inspection. The tyrosinase minigene was expressed properly in the different pigmented cell types within the eye. These results confirm the hypothesis that albinism is due to loss of endogenous tyrosinase activity. Through genomic sequencing studies, the albino mutation in laboratory mice was subsequently identified as a single–base pair change that causes a conversion from cysteine to serine at amino acid 103 of the tyrosinase protein. The tyrosinase minigene is a useful new research tool, since it allows transgenic mice to be recognized by simple visual inspection.

The other major focus of our research is the characterization of recessive mutations that are caused by integration of transgenic DNA. Such mutations are valuable, because the transgenic DNA can be used to isolate the inactivated genes. Once the transgenic DNA has integrated into the genome, it becomes linked to adjacent genomic sequences. Using recombinant DNA technology, we can fragment the genome into a large number of segments, and the segments can be individually cloned to give a genomic "library." The library can be screened by hybridization techniques to identify the clones that contain the transgenic DNA along with the adjacent genomic sequences. In this fashion, the transgenic DNA provides a molecular tag that can be used to isolate genes that play essential roles in normal mammalian development.

In one of our mouse families, the transgenic insert has inactivated the *downless* gene, a gene essential for normal induction of hair follicle formation. In another transgenic family, there is a mutation that blocks sperm development. For both of these mutants, the genomic sequences that flank the transgenic DNA have been isolated. Putative coding regions have been identified and sequenced, and efforts are in progress to confirm that the proteins encoded by these regions play essential roles in specific stages of development.

Such studies can provide important insights into the mechanisms that specify differentiation and morphogenesis during mammalian embryonic development, and may provide clues for strategies to help prevent birth defects in humans.

Structural Studies of DNA-binding Proteins



Carl O. Pabo, Ph.D.—Associate Investigator

Dr. Pabo is also Professor in the Departments of Molecular Biology and Genetics and of Biophysics at the Johns Hopkins University School of Medicine. 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. In July 1991, Dr. Pabo will move to the Massachusetts Institute of Technology as Professor in the Department of Biology.

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 year, 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 has recently solved this structure. The overall arrangement of the helix-turn-helix unit and the amino-terminal arm are similar to the arrangement seen in the *engrailed* complex, but there are a number of differences in the critical side chains used for recognition. There also is a rich background of biochemical and genetic data about α2 that should help us understand the precise role of this homeodomain in recognition and regulation.

Structure of a Zinc Finger–DNA Complex

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, at the Johns Hopkins University), he cloned and expressed a three-finger peptide and crystallized the peptide-DNA complex. The zinc fingers recognize 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 fingers are used in a modular fashion, they may be the ideal motif to use as we try to design novel DNA-binding proteins.

Tools for Analyzing Protein-DNA Interactions

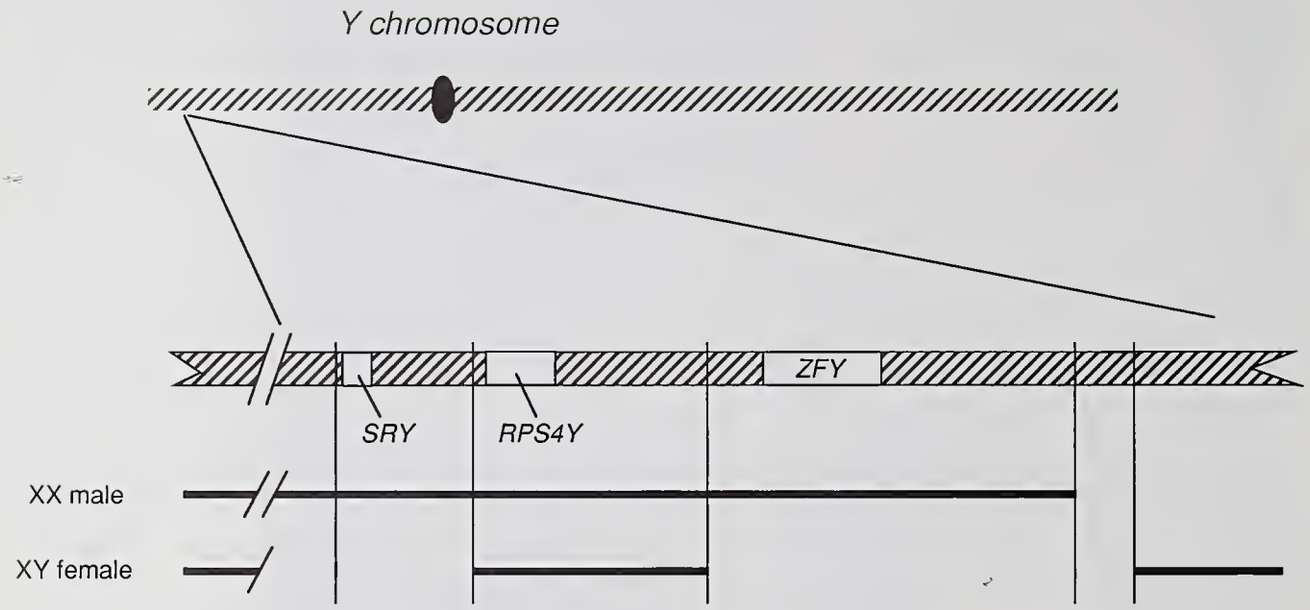
We are working on several other projects that should help to develop the background for a systematic analysis of protein-DNA interactions.

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.

We have also been developing testing strategies for computer-aided protein design. Our programs can systematically consider a large number of sequences and conformations, and we are using these programs as we attempt to design zinc finger proteins that will recognize novel binding sites. We also are developing genetic strategies for selecting zinc finger proteins that recognize desired target sequences. We hope that this combination of structural analysis, computer-aided protein design, and genetic selection will give us a better understanding of protein-DNA recognition and allow us to design zinc finger proteins that recognize novel target sites.

A representation of the structure of the engrailed *homeodomain-DNA complex as determined by x-ray crystallography. The protein (shown in red and yellow) binds to the DNA (blue) and makes numerous contacts with a 4–base pair region (purple) of the double-helical DNA. One α-helix from the homeodomain (which looks like a red circle in projection) fits into the major groove of the DNA, and an extended region of the protein backbone (which appears as a red ribbon in the figure) fits into the minor groove of the DNA.*

This structure was determined by Charles Kissinger and Carl Pabo. Illustration originally printed as cover of Cell *63(3), 1990. See also Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B., and Pabo, C.O. 1990.* Cell *63:579–590.*



The tiny region of the human Y chromosome implicated in both sex determination and Turner syndrome is shown in expanded view below the schematic diagram of the entire Y chromosome (the black oval represents the centromere of the Y chromosome). The region contains the genes SRY, RPS4Y, *and* ZFY. *The black lines at the bottom of the diagram depict the portion of the Y present in two pivotal patients, one an XX male and the other an XY female (lacking features of Turner syndrome).*

Research of David C. 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 and Assistant Professor of Biology at the Massachusetts Institute of 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 vary 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 variability, it is important to distinguish between variation in a single gene, called Mendelian, and massive multigene variation, called chromosomal.

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 genetic variability 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 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 have 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 patterns of 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 small region 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 a few years ago, and *SRY*, a gene recently described by British scientists.

Both *ZFY* and *SRY* appear to encode DNA-binding proteins that are likely to regulate the transcriptional activity of particular but unknown target genes. The evidence is strong that *SRY* is a sex-determining gene. Much less clear is the role, if any, that *ZFY* plays in the process. We hope to learn more about the function of the *ZFY* gene by simultaneously analyzing a closely related gene, *ZFX*, that we identified on the X chromosome.

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. 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 at least 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 very existence of related but nonidentical ribosomal protein genes on the X and Y chromosomes suggests that the ribosomes of human males may differ slightly from those of females. It will be a surprise to many if the differences between the sexes extend all the way down to the most fundamental and vital of intracellular machines!

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.

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 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 at 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, 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 gene's expression in the 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 to hepatocytes, and protamine I to male germ cells. More recently, we have begun to locate the elements involved in directing the expression of genes in the catecholamine biosynthesis pathway to specific neurons.

Because the regulatory elements from one gene can often be used to control another, the expression of many interesting genes can be directed to a particular cell type and the consequences on cellular development and function 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 SV40 virus, 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 the 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 are more aggressive than those in mice carrying a single gene, suggesting that these genes act cooperatively.

We have recently shown that the expression of genes 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. HBV infects millions of people worldwide, and the incidence of liver cancer among them is high. In transgenic mice, expression of this gene resulted in synthesis of the viral surface antigen and envelope protein, which aggregated within the secretory apparatus of the liver cells, causing cellular injury and death. When the mice were more than a year old, they developed liver cancer. In a similar case, expression in liver of plasminogen activator, a protease, results in selection of cells

that inactivate the protease-encoding transgene. These cells grow and repopulate the liver, and the mice 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 a toxic environment of liver injury results in genetic damage that predisposes the cells to malignant transformation.

It is also possible to develop transgenic mice that mimic some human genetic diseases. For example, we have recently 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 the β-globin gene), we have generated mice that produce as much human hemoglobin as mouse hemoglobin. When the mutant β-globin gene from people with sickle cell 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 catecholamine neurotransmitters: dopamine, norepinephrine, and epinephrine. These genes are expressed in certain neurons of the central nervous system, in the peripheral nervous system, and in the adrenal medulla. The control elements from these genes are being tested in conjunction with reporter genes whose products can be easily visualized to assess when and where they are expressed during development. Subsequently, we intend to use the control elements to direct the expression of other genes to these neurons with the aim of affecting the decisions they make during the process of forming functional connections with target cells.

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. 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 essential roles in the body's ability to respond to stress. Two different parts of the gland, an inner medulla and an outer cortex, produce discrete agents 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 synthesized by 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. Certain of these enzymes, such as the cholesterol side-chain cleavage enzyme (SCC), are expressed in all steroidogenic tissues. In contrast, steroid 11β-hydroxylase (11β-OHase) and 21-hydroxylase (21-OHase) are only expressed in the adrenal cortex and largely determine the unique ability of this tissue to make glucocorticoids and mineralocorticoids. The physiological regulators of these two classes of adrenal steroids differ markedly despite the shared role of the 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, termed the promoter region, contains most sequences important in transcriptional regulation of other genes. We first showed that the 5′-flanking regions of the steroidogenic enzymes retained all information required for adrenal-selective and hormonally inducible gene expression.

Next, in a combination of structural and functional studies, we identified a protein, termed steroidogenic regulatory protein, that played a major role in regulating the expression of the steroidogenic enzymes. This protein was only present in nuclear extracts from steroidogenic cell lines, suggesting that it contributed to the cell-selective expression of these genes. Moreover, steroidogenic regulatory protein appeared to interact with the promoter regions of all three steroidogenic enzymes, suggesting that it coordinates the expression of this entire network of enzymes in adrenal cells.

Now that this major regulatory protein has been identified, our next goal is to understand how it affects the expression of 21-OHase, 11β-OHase, and SCC. These studies require sufficient amounts of protein for biochemical analysis, and we have therefore initiated studies using cow adrenal glands as a source of protein. These experiments demonstrated that bovine protein behaved very similarly to steroidogenic regulatory protein isolated from mouse adrenocortical tumor cells. By sequential purification steps, we have markedly enriched the bovine protein to a degree that should permit determination of its amino acid sequence.

Attempts are also under way to raise antibodies against the partially purified bovine protein. The combination of specific antibodies and amino acid sequence data should allow us to clone the gene encoding this key regulator. By comparing the primary structure of steroidogenic regulatory protein with that of previously described transcriptional regulatory proteins, such as those of the steroid hormone receptor, we may gain new insights into the mechanisms that regulate the adrenal steroidogenic enzymes. The availability of specific probes and antibodies to steroidogenic regulatory protein 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. They may also increase our understanding of the general mechanisms of tissue-specific gene expression.

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. As such, it 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. Furthermore, in certain animal models with a more carefully defined genetic basis, the inheritance of hypertension has been clearly linked to differences in activity of 11β-OHase, one of the steroidogenic enzymes described above. We are investigating the function of this protein in more detail.

Initial studies implicated a single 11β-OHase protein in the biosynthesis of both mineralocorticoids and glucocorticoids. We now have definitive evidence that the mouse has two 11β-OHase genes, one of which is expressed at significantly higher levels than the other. By analogy to the rat, for which biochemical characterization has been more extensive, it is likely that the gene expressed at higher levels participates in the formation of glucocorticoids in the inner zone, whereas the other gene, encoding aldosterone synthetase, produces mineralocorticoids in the outer zone.

Although the two genes are very similar in their coding regions, which specify the proteins' composition, they differ markedly in their 5′-flanking regions, which regulate the proteins' production. We are now seeking to define the precise role of these two genes in the production of steroid hormones. In particular, we will compare and contrast the promoter regions of the two genes to identify both shared elements and unique elements responsible for zone-specific expression.

Based on our preliminary results, we plan to use these promoter regions in transgenic mice to target gene expression specifically to either the outer or inner 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 the hypertensive state.

Next, we want to express the aldosterone synthetase gene in the inner, glucocorticoid-producing zone. As discussed above, this gene normally produces mineralocorticoids in the outer zone. We have prepared a hybrid gene with the inner-zone-specific promoter driving expression of the aldosterone synthetase gene. We anticipate that this hybrid gene will synthesize large amounts of mineralocorticoids, thus creating a genetic form of hypertension. Moreover, treatment with glucocorticoids, which will suppress expression of the hybrid gene in the inner zone, should alleviate the hypertension. This model may therefore mimic a subset of human patients with glucocorticoid-responsive 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 secondary education in Mexico and Switzerland and 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 is currently studying the interactions at the molecular level between the nervous and immune systems. These analyses are focused on the roles of certain receptors in cellular differentiation and signal transduction and, more recently, on the organization of other unique cell surface components. Projects that are under way involve the expression of tachykinin receptors in cloned lymphocyte, epithelial, and neuronal cell lines and an analysis of their signaling pathways. In addition, we are studying the molecular biology of synaptogenesis, in particular the function of the molecule agrin.

With Richard Scheller (HHMI) and Fabio Rupp at Stanford University, I have worked on the cloning and structural analysis of the agrin molecule from the rat. Agrin belongs to a family of extracellular synaptic organizing molecules that have been shown to cause the aggregation of acetylcholine receptors and acetylcholinesterase on the surface of regenerating and developing muscle fibers. A cDNA library constructed from poly(A)$^+$ RNA of embryonic rat spinal cord was screened, and a number of inserts were isolated that predicted amino acid sequences similar to those of agrin identified in the *Torpedo* electric organ.

A number of these clones have now been sequenced. We have constructed fusion proteins in order to raise specific antibodies, and we are studying the expression of the message for agrin in various tissues during the development of the rat. The deduced sequence contains several epidermal growth factor (EGF)-like domains and regions similar to the trypsin inhibitor enzyme. Moreover, Northern blot analysis reveals a message of about 8.2 kb in embryonic brain and spinal cord, which in the adult decreases about 10-fold and is only detectable in the cord. Current efforts are focused on expressing agrin in a number of cell lines and studying its receptor-clustering activity. In the future we will analyze its functionally active domains, possible binding sites, and cellular distribution in embryonic and adult rat tissues.

With information derived from the cloning of the rat substance P (SP) receptor (SPR), my laboratory is studying the expression and the signaling properties of the SPR in a number of cell lines. We have demonstrated that the SPR signals simultaneously by both the cAMP and IP_3 (inositol 1,4,5,-trisphosphate) pathways.

With J. Sudduth-Klinger and C. Christian, I have transfected the SPR cDNA into Jurkat lymphocytes and PC12 cells and established stable clones that express functional SPR. The initial characterization of the lines has been carried out. Jurkat-SPR cells will be studied to see if, when stimulated by SP, they alter the expression/production of specific cytokines or their receptors. Moreover, the coordinate regulation of CD_{10} (the cell surface neutral endopeptidase that deactivates SP) in Jurkat-SPR will be examined. We will study whether CD_{10} activation or inactivation by thiorphan (a neutral endopeptidase inhibitor) changes SP responses. In addition, we will also study whether CD_{10} coprecipitates with the SPR. Preliminary immunohistochemical studies from rat lung and gut suggest that the SPR and CD_{10} are colocalized. These experiments will for the first time allow us to elucidate the effects of SP on immune responses at a unicellular level in a well-defined system.

I will study with Mark Gilbert how the SPR influences PC12 cells before and after differentiation with nerve growth factor (NGF). Functional studies are under way to examine whether SPR expression alters growth characteristics and morphology of the cells. Experiments from other laboratories done with primary cultures have shown that ^{125}I-SP binding is located in the growth cones, suggesting the receptor may play a role in certain aspects of neuronal development.

Molecular Basis of Lymphocyte Signaling

Roger M. Perlmutter, M.D., Ph.D.—Associate 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 the 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.

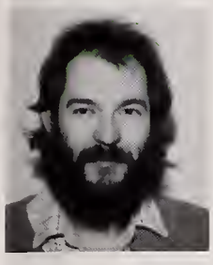
Initially we identified a lymphocyte-specific enzyme that is similar in structure to proteins known to transmit growth-promoting signals in nonlymphoid cells. This protein kinase modifies the behavior of target proteins by catalyzing the addition of phosphate groups onto specific tyrosine amino acids in substrate proteins. The gene, *lck*, 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 the *lck* gene was sufficient to unmask its ability to deregulate the growth of cells grown in tissue 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 augmented expression of a protein kinase encoded by the *fyn* gene produced lymphocytes that manifest 10-fold more vigorous responses to antigen stimulation than do normal cells. Interference with this *fyn*-controlled pathway yields cells that are refractory to any antigenic challenge.

Using these and other experimental strategies, we hope to deduce eventually a "wiring diagram" that will adequately describe signal transduction mechanisms employed by lymphocyte antigen receptors. 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.

Signal Transduction and the Specification of Cell Fates



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.*

CONSIDERING the number of cells in an organism and the seemingly infinite possibilities for developmental errors, it is amazing that individuals retain their characteristic shapes. Cells of specific tissues interact at multiple stages of development, and their communication is vital to the establishment of patterns. Unraveling the mysteries of the intercellular communication processes is the interest of our laboratory.

We are studying these processes in *Drosophila*, since mutations are easy to generate, known mutant phenotypes are plentiful, and techniques for cloning and transferring genes are well established. Our goal is to identify and characterize, using genetic and molecular techniques, the components of specific signal transduction pathways implicated in the control of pattern formation.

Cell fate along the anterior-posterior egg axis is determined by three groups of maternally expressed genes: those of the anterior, posterior, and terminal class systems. We are investigating the terminal class system, which controls cell fate at both ends of the *Drosophila* embryo. In embryos lacking maternal terminal class functions, both head and tail structures are deleted.

The activities generated by the terminal class system are believed to be propagated through a signal transduction mechanism that involves a phosphorylation cascade. The current model is that the putative transmembrane tyrosine kinase receptor, encoded by the gene *torso*, is activated only at the egg termini. This localized activation of *torso* then triggers a phosphorylation cascade that culminates in transcription of the zygotic terminal gap genes *tailless* and *huckebein*, which are known to encode transcription factors.

Our molecular and genetic analysis of the terminal class system focuses on the proteins involved in transducing the signal from *torso*'s membrane-bound kinase to the nucleus, and thus far we have characterized two genes, *l(1)pole hole* and *l(1)corkscrew*. We have shown by genetic epistasis experiments that both act downstream of the *torso* protein activity. The molecular characterization of *l(1)pole hole* has strengthened the current model for transfer of the *torso* signal, since the *l(1)pole hole* gene product is the homologue of the mammalian *Raf-1* proto-oncogene, which encodes a serine/threonine kinase. Molecular characterization of *l(1)corkscrew*, currently in progress, will provide more insight into this signal transduction pathway.

To identify other molecules involved in the *l(1)pole hole* signal pathway, we have undertaken two genetic approaches to screen for second-site suppressors and enhancers. In the first type of screen, we have used a mutation in the *l(1)pole hole* gene that reduces its level of activity. Progeny with this gene are not viable. They can live, however, in the presence of a second-site suppressor. We have conducted extensive genetic screens to search for these suppressors and have recovered a number of them. Presumably, these suppressors identify proteins involved in the *l(1)pole hole* signal transduction pathway that are able to increase or bypass the *l(1)pole hole* activity. Future work will encompass a detailed characterization of these suppressors.

In the second type of screen, we will search for second-site modifiers that affect the activity level of an *l(1)pole hole* gain-of-function mutation. Since such hyperactive gain-of-function *l(1)pole hole* mutations have not been recovered by conventional genetic techniques (presumably they are lethal to the fly), we decided to "switch on" an activated form of *l(1)pole hole* in nonvital organs of the adult fly to generate a visible dominant phenotype. Using the activation properties of the yeast GAL4 protein, which activates only those genes bearing a GAL4-binding site within their promoters, we were able to express a hyperactive *l(1)pole hole*–modified protein in the fly's eye imaginal disc. We are using the resultant dominant eye phenotype to screen for mutations that suppress the phenotype. Presumably the suppressors will identify proteins that can reduce the level of *l(1)pole hole* activity.

We believe that such experiments will ultimately permit genetic isolation of more components required in the *l(1)pole hole* signal transduction pathway. Although our primary objective is to understand the role these molecules play in

Drosophila, our work may be directly relevant to an understanding of development in other organisms as well. The high levels of homology between *torso* and the gene for platelet-derived growth factor (PDGF) and between *l(1)pole hole* and *Raf-1* suggest that molecules interacting with *l(1)pole hole* might be conserved in the *Raf-1* pathway of higher vertebrates.

Gene Regulation and Immunodeficiency

B. Matija Peterlin, M.D.—Assistant 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 the 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 research interest in this area. 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 (MHC) 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 or fail to make antibodies or have autoimmune diseases. In addition, this autosomal recessive syndrome is one of the two known inherited deficiencies of a regulatory gene in humans.

By fusing in tissue culture defective cells from different patients and those obtained by mutagenesis, 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 might 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 encoding 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. The remaining two cDNAs encode a B cell–specific helix-loop-helix protein, NF-IL6, and an ETS-like protein. By expressing one of these full-length cDNAs in human cells, we hope to rescue class II gene expression in one type of BLS. Besides direct biochemical studies, we are also using genetic approaches to rescue regulatory defects in this disease.

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. Factors assembled near the site of initiation of HIV transcription bring 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, and 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. 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 transcription 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 MHC 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. 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.

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 effecting 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. The model system used in our studies employs the A431 line of human epidermal carcinoma cells, which are highly responsive to EGF.

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 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. Preliminary data suggest that in membranes and whole cells this activity is regulated via phosphorylation, and a protein kinase has been identified that phosphorylates and activates the phosphatidylinositol kinase *in vitro*. Work is under way to characterize this event.

Another enzyme involved in phosphatidylinositol metabolism is a phosphatidylinositol monophosphate phosphatase. This enzyme catalyzes the reverse of the reaction catalyzed by the phosphatidylinositol kinase—that is, it removes the phosphate from the 4 position of the inositol ring on phosphatidylinositol monophosphate. Although this enzyme has not been studied previously, its position in the metabolic pathway of phosphatidylinositol suggests it may be important in the overall regulation of the pathway. We have identified its activity in A431 cells and 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 140,000-Da glycoprotein. Unlike other enzymes in this pathway, this enzyme appears to have an extracellular domain. The possibility that the activity of this enzyme 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 are in progress to determine whether EGF-stimulated changes in cell shape can be attributed to the generation of novel inositol phospholipids or is due to the ability of the EGF receptor to catalyze the phosphorylation of cytoskeletal proteins.

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, as 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. Desensitized EGF receptors cannot undergo this EGF-induced dimerization. As a result, they cannot transduce the hormonal signal to the interior of the cell.

Evidence suggests 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 plan to alter this site and study the effect of this mutation on the function of the EGF receptor.

The structure of adenosine deaminase complexed with 6-hydroxyl-1,6-dihydropurine ribonucleoside, a nearly ideal analogue of the transition state of catalysis, and the zinc cofactor, as determined by crystallographic technique. Adenosine deaminase plays a central role in maintaining immune competence; impairment of its catalytic activity is associated with severe combined immunodeficiency (SCID), affecting both T and B lymphocytes.

Research of David K. Wilson and Florante A. Quiocho.

Protein Structures, Molecular Recognitions, and Functions

Florante A. Quiocho, Ph.D.—Investigator

Dr. Quiocho is also Professor of Biochemistry 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 joined the Rice University faculty as an assistant professor of biochemistry. Leaving Rice as a full professor, he joined the faculty of Baylor. 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.

FORMATION of complexes between proteins and their ligands is the basis of biological specificity and activity. Our long-term goal is to elucidate the atomic interactions between proteins and ligands in a variety of biological systems. While x-ray crystallography is our primary approach, we also do correlative studies employing biochemical and recombinant DNA techniques.

Adenosine Deaminase

Adenosine deaminase (ADA), present in virtually all mammalian cells, catalyzes the irreversible deamination of adenosine to inosine. The enzyme has a central role in the normal development of the immune response, and its genetic absence leads to severe combined immunodeficiency disease (SCID). It is also specifically involved, or its levels are changed, in a variety of diseases, including AIDS (acquired immune deficiency syndrome), anemia, and various lymphomas and leukemias. It also appears to modulate synaptic transmission.

We have previously obtained crystals of ADA in the presence of purine ribonucleoside, a substrate analogue. Employing x-ray crystallographic techniques, we have recently determined and refined the three-dimensional structure of ADA at 2.4 Å resolution. The deaminase has a parallel $(\alpha\beta)_8$ barrel structure. In the course of the structure refinement, we discovered that ADA contains a zinc cofactor and that the bound ligand is 6-hydroxyl-1,6-dihydropurine ribonucleoside (HDPR), a nearly ideal transition-state analogue, instead of the substrate analogue. The zinc and HDPR are bound and sequestered in a deep pocket located at the carboxyl-terminal end of the β barrel.

The potent transition-state analogue (inhibition constant estimated at 10^{-13} M) is held tightly in place by the coordination of the 6-hydroxyl to the zinc ion and by the formation of nine hydrogen bonds with seven residues. The enzyme preferentially recognizes the 6R distereomer of the ribonucleoside.

The discovery of the bound zinc and transition-state analogue has enabled us to understand in great detail the mode of action of ADA. We have also been able to provide a molecular explanation of the mutations of the enzyme that result in SCID. Inherited deficiency of ADA accounts for about one-third of SCID cases. The sequences of eight point mutations of ADA from five cell lines derived from affected patients have been reported. All but two of these mutations occur close to an active-site pocket or to peptide segments that line it. Consequently, the deleterious effects of the six mutations can be ascribed to structural changes that affect the active-site geometry.

Knowledge of the ADA structure paves the way for determining the structures of mutants, some mimicking those related to deaminase deficiency, and other complexes with various analogues and inhibitors. We have also obtained crystals of the deaminase in the presence of two other analogues (one currently being used for chemotherapy in certain leukemias) and seven compounds with various therapeutic uses, such as sedatives, anxiolytics, analgesics, and relaxants.

The High Specificity of a Phosphate Transport System

Phosphorus in the form of orthophosphate is probably one of the most essential nutrients for cell life. The transport of phosphate into cells and organelles (mitochondria, chloroplasts, etc.) is highly specific. Obtaining the extremely well-refined 1.7-Å structure of the liganded form of phosphate-binding protein has revealed for the first time the atomic features responsible for the exquisite high selectivity for phosphate and the exclusion of sulfate. The phosphate, which is totally dehydrated and buried in the cleft between the two domains of the protein, is held in place by a total of 12 hydrogen bonds formed with protein groups, one of which is the carboxylate group of an Asp residue. This residue is responsible for the recognition of a proton on the phosphate.

Structure and Recognition of Epitopes

Knowledge of the active conformation of a dominant neutralizing epitope (antigenic determinant group) is essential to understanding the molecular basis of the immunogenicity of the segments and to obtaining possible leads for vaccine design. Recombinant technology is being used to insert, without loss of immunogenicity, foreign epitopes, especially of proteins with unknown structures, onto the surface of the maltodextrin-binding protein of *Escherichia coli*, in sites known to be permissive to changes. The structure of the protein containing such an epitope provides information about the epitope's configuration and how its immunogenicity is preserved. Currently under investigation are epitopes of type 1 poliovirus and HIV (human immunodeficiency virus) gp120, the coat protein of the AIDS virus.

Antibody-Antigen Interaction

We have previously reported the crystallization of the Fab fragment of a monoclonal antibody raised against a dominant neutralizing determinant of gp120. Diffraction data from these crystals, out to 2.8 Å resolution, have been collected by means of a synchrotron facility. Attempts to determine the structure, using molecular replacement techniques, are under way. X-ray structure determination should reveal details of molecular recognition of an antigen and important facets of gp120 that are amenable to antibody intervention.

We have determined the structure of the Fab fragment of an antibody against cell surface polysaccharide antigen of *Shigella flexneri* in complex with tri- and pentasaccharide-containing determinants of the O-antigen serotype of the bacteria. Determination of the structures of these crystals is in progress.

Molecular Recognition of Carbohydrates, Charged Substrates, and Calcium

The family of binding proteins that serve as initial receptors for bacterial active transport and chemotaxis is an excellent system in which to study the molecular recognition of a variety of ligands and structure-function relationships. To date, we have determined and refined at better than 1.9 Å resolution the structures of seven different periplasmic receptor proteins: those binding L-arabinose, D-galactose/D-glucose, maltodextrin, leucine/isoleucine/valine, leucine alone, sulfate, and phosphate. Using site-directed mutagenesis, coupled with analysis of substrate binding to mutant proteins by crystallographic and other techniques, we are beginning to dissect various aspects of the molecular recognition of carbohydrates, charged ligands such as sulfate and phosphate, and calcium ion.

A method different from site-directed mutagenesis for probing the contribution of specific hydrogen bonds in protein-carbohydrate interactions involves specific modifications of the ligand. The binding of deoxy and fluorodeoxy sugar analogues to the arabinose-binding protein has been analyzed by equilibrium and very high resolution crystallographic techniques. The complexes have structures virtually identical to the wild type and reflect more realistically the hydrogen bond interaction being probed.

Molecular Mechanisms of 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 that 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 adenylyl cyclase in olfactory neurons suggested an analogy to visual signal transduction. Moreover, activation of adenylyl cyclase in preparations of rat olfactory cilia depends on the presence of guanine nucleotides. The observation of odorant-stimulated GTP-dependent adenylate cyclase activity argues strongly for a GTP-binding protein–coupled signal transductory pathway. We have identified a GTP-binding protein expressed exclusively in olfactory sensory neurons. Moreover, this G protein, G_{olf}, is 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 olfactory receptors. We have demonstrated that G_{olf} can couple receptor activation to increases in intracellular cAMP in cell lines deficient for G_{sa}.

Does the identification of an olfactory-specific G protein provide any insight into the nature of the olfactory receptor? A number of G protein–coupled receptors have been cloned and characterized; each consists of a single glycosylated polypeptide chain containing seven membrane crossing regions. There is limited but significant homology among members of this family, and we are presently attempting the direct cloning of the olfactory receptors based on these similarities. We have recently identified several novel G protein–coupled receptors expressed in olfactory epithelium. Their role in olfaction remains unclear. Identification of genes encoding olfactory receptors may reveal how receptors are able to detect thousands of different odors. Specifically, does each of the several million olfactory sensory cells encode a distinct receptor protein?

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 en-

zymes 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. It appears that one form of this enzyme is expressed by olfactory sensory neurons. This 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.

We have recently identified a number of cDNA clones that encode proteins expressed only in the mature sensory neurons and are presently 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 have also initiated a project to generate cell lines derived from cells of olfactory neuronal lineage. Transgenic mice have been identified that carry the SV40 T antigen oncogene linked to a transcriptional promoter for an abundant olfactory neuron-specific gene product. These animals develop tumors originating in the olfactory neuroepithelium. From these tumors, we have established cell lines in culture that express a number of characteristic neuronal markers. These cell lines express several of the neuron-specific markers we previously identified. This approach may be generally applicable to the isolation of defined cell lines in many other systems.

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.—Assistant 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 studying 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 antibodies produce a sudden and devastating 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 so as to study its structure. Because the collagen 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.

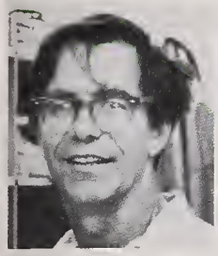
In collaboration with Billy Hudson (Kansas City), we used knowledge of the primary structure to identify several potential antibody-binding sites (epitopes) in the $\alpha 3$ molecule. We synthesized short peptides and used them to test the binding of some of these sites, which we localized to within a small region of 12 amino acids. 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.

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. 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.

Jaap IJdo, a Howard Hughes associate in our laboratory, has been interested in the structure of telomeres, the ends of chromosomes. He has shown that banks of repetitive sequence reminiscent of human telomeres (TTAGGG) are also present at other sites within the human genome. One of the most interesting of these banks 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. Dr. IJdo has cloned this point and shown that it consists of a head-to-head telomere-telomere fusion. He is investigating the possibility that this point of fusion is also responsible for a rare form of chromosome fragility that has been observed to occur at or very near this site.

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 Dr. Reichardt's 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, 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—e.g., laminin and N-cadherin—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.

Work on trophic factors in the past year has focused on those related to nerve growth factor. The human homologue of one of these, brain-derived neurotropic factor, was isolated, and NT-3, a novel trophic factor in the nerve growth factor family, was identified. Each acts on distinct but partially overlapping sets of neurons.

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 nerve growth factor–related trophic factors is the low-affinity nerve growth factor receptor. Antibodies to this protein have been prepared and used to show that it is distinct from a second protein (or set of proteins with similar sizes) that constitutes a distinct high-affinity receptor for nerve growth factor. The same antibodies have been used to show that nerve growth factor–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 nerve growth factor.

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 among these isoforms, making it possible for cells to exhibit different responses to individual isoforms.

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 ex-

hibiting 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 appears to modulate the signals conveyed by integrin receptor binding to laminin and other proteins. These signals can be modulated by trophic factors, such as nerve growth factor.

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 on central and peripheral cells. 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. The cell adhesion molecules include members of the immunoglobulin superfamily and members of a new family of calcium-dependent adhesion molecules, the cadherins.

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. 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. 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.

T Lymphocyte Biology and Genetics

Robert R. Rich, M.D.—Investigator

Dr. Rich is also Professor of Microbiology and Immunology and of Medicine at Baylor College of Medicine and Attending Physician at Baylor College of Medicine and affiliated hospitals. He received his undergraduate education at Oberlin College and his M.D. degree from the University of Kansas. After a residency in internal medicine at the University of Washington, he had postdoctoral fellowships at the NIH with Sheldon Wolff and Charles Kirkpatrick and at Harvard Medical School with Baruj Benacerraf and Carl Pierce.

MULTICELLULAR organisms have evolved complex mechanisms to differentiate themselves from their environments and to perceive and defend against potential threats to their integrity. The immune system of vertebrates is a sophisticated defense system organized to deal with attack at the microbial and/or molecular level. Recent data indicate that microbial organisms similarly have evolved mechanisms to enhance the effectiveness of their interactions with specific recognition elements of vertebrate immune systems. The primary focus of specific recognition by the immune system is a specialized subpopulation of lymphocytes, termed T cells, that is responsible for discriminating molecules perceived as self from those that are foreign. Several families of genes are directly involved in this process, including genes that encode the T cell receptor for antigen and those of the major histocompatibility complex (MHC).

The MHC genes, which exist as a linked cluster on a single chromosome, are intimately involved in the definition of "self" for cells of the immune system. The importance of the MHC in self-nonself discrimination was appreciated long before its essential role in antigen recognition was defined. As suggested by its name, the MHC constitutes the major barrier to exchange of tissue grafts between members of a vertebrate species.

T cell receptors recognize foreign antigen solely in the context of self MHC molecules, which are expressed on the surfaces of cells specialized for antigen presentation. Recent data suggest that recognition is accomplished by the specific binding of small peptide antigens within a groove on the outer face of MHC molecules. T cell receptors are thought to be selected during differentiation in the thymus for their capacity to bind this binary complex of foreign peptide and self MHC.

Our objectives include definition of the genes and gene products of the MHC in mice and humans and incorporation of such information into an understanding of T cell recognition of foreignness. The relationships between the structure of MHC molecules and their biological functions are of particular interest. Two major classes of MHC molecules have been defined by their distinctive structural features. Molecules of both classes are important in self-nonself discrimination.

T cells involved in induction of immune responses primarily employ class II MHC molecules. Three major types of human class II molecules are recognized: DP, DQ, and DR. Previously we employed cloned T cells to define the contributions of each type to the recognition of foreign antigens and to investigate the cellular mechanisms involved in antigen presentation. A DR molecule was identified with unique biochemical features and distinctive defects in its capacity to function in antigen presentation. This defect was highly associated with a deficiency in activity of linked genes controlling an enzyme important in the biosynthesis of adrenal steroids. Because an understanding of the DR molecule should provide insight into general structure-function relationships of MHC molecules, we sequenced cDNAs encoding its constituent polypeptides. The biochemical and functional anomalies were associated with point mutations in two adjacent codons, resulting in amino acid substitutions within the β-chain of the DR molecule. These substitutions are hypothesized to affect interactions between the β- and α-chain polypeptides and may affect binding of foreign antigenic peptides. Studies in progress are designed to elucidate the molecular mechanisms accounting for the unique properties of this molecule.

Our studies of MHC class II–dependent T cell stimulation led us to analysis of the mitogenic activities of a family of staphylococcal exotoxins that have been of considerable medical interest as the etiologic agents of staphylococcal food poisoning and toxic shock syndrome. The recent appreciation of unique aspects of the interaction of these bacterial products with T cell receptors, and their capacity to thus shape the T cell repertoire, has led to their designation as "superantigens" (Philippa Marrack [HHMI] and John Kappler [HHMI], National Jewish Center for Immunology and Respiratory Medicine, Denver).

We initially reported that MHC class II molecules are the specific cellular receptors for a prototype of this family of molecules, staphylococcal enterotoxin A (SEA).

Interest in the structural basis of superantigen activities led us to assess their binding affinities to MHC class II molecules and the effects of biochemical modifications on binding and function. We observed a direct correlation between binding affinity as a linear function and the logarithm of the effective dose for T cell stimulation. Thus relatively small changes in binding affinity had marked effects on enterotoxin potency. A common feature of members of the enterotoxin family is a small intrachain disulfide loop in the central region of the molecule. We assessed the role of this loop and adjacent amino acids on functional activity and demonstrated that the disulfide bond was essential to T cell mitogenic activity but had little effect on MHC class II binding or monocyte stimulation. We concluded that the inability of these modified toxins to induce T cell proliferation reflected their failure to interact effectively with the T cell receptor for antigen rather than with class II molecules.

For several years we have also studied "nonclassical" MHC class I genes and antigens of the mouse. These genes, which constitute the majority of class I genes, were termed nonclassical because they had not been thought to be involved in conventional antigen presentation and self-nonself discrimination. We have been specifically interested in the maternally transmitted antigen (Mta). Discovered by Kirsten Fischer Lindahl (HHMI, University of Texas Southwestern Medical Center at Dallas), Mta is recognized by cytolytic T cells and is unique in its mode of inheritance. Offspring of mice with different forms of the antigen always express the maternal form. Both nuclear and mitochondrial genes are involved in Mta expression, but the maternally inherited polymorphism is determined by a mitochondrial gene product. This is a hydrophobic peptide encoded at the amino terminus of the *ND1* mitochondrial gene. We have investigated structural properties of this peptide that enable it to bind and be presented by a specific nonclassical class I gene product, Hmt. The essential feature of this peptide is a biochemical substitution (an amino-terminal formyl residue) that is characteristic of the first amino acid of mitochondrial proteins but is not found on proteins synthesized from nuclear genes. Proteins synthesized within mitochondria share this property with proteins synthesized by bacteria. We therefore investigated the capacity of certain bacterial peptides to bind similarly to the Hmt class I molecule. Three of the first six synthetic bacterial peptides characterized displayed this property, but if and only if they also had the terminal formyl residue. Unmodified peptides or peptides modified with other biochemical radicals were unable to bind.

These data suggest that the Hmt class I molecule may represent a specialized host defense mechanism for recognition of proteins of bacterial origin. As such they may provide insight into the biological functions of the nonclassical MHC class I antigens. Additionally, they offer the first evidence of true biochemical specialization in the binding of antigenic peptides to a specific MHC molecule.

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*, which is amenable to behavioral as well as biochemical and genetic approaches.

A major objective in the field of RNA processing is to understand the rules that govern pre-mRNA splicing. It is now more than 10 years since splicing was discovered in mammalian cells and viruses. This discovery revolutionized the field of genetics, since it refuted a major tenet—that genes are intact units. It became clear that genes are divided into segments and that the incoherent material (the introns) is discarded, while the retained sense material needed to code for proteins (the exons) is carefully sewn together. Neither the efficiency nor the specificity of the process is well understood.

More recently it emerged that all eukaryotes (including *S. cerevisiae*) undergo the same splicing process with essentially the same biochemical machinery. This includes a myriad of protein factors, most of which remain uncharacterized if not unidentified, and five small nuclear ribonucleoprotein particles (snRNPs). Each snRNP contains one or two small molecules of RNA and perhaps 5–10 proteins. The pre-mRNA substrate and many of these factors, including the five snRNPs, assemble into the spliceosome, a large particle within which the cleavage and ligation events of splicing take place.

We are particularly interested in the early events of the spliceosome assembly process, as it seems to be related to those of intron recognition or definition. The highly ordered assembly process must involve recognition of an appropriate pre-mRNA substrate to initiate the splicing process and prevent unproductive transport of the unspliced pre-mRNA to the cytoplasm. In addition, these early events are likely related to the mysterious prevention of exon shuffling—i.e., exons are usually sewn together in the correct order.

We have focused on U1 snRNP, as there was reason to believe that it features prominently in recognition of one of the two intron splice sites (the 5′ splice site) and even in catalytic specificity. Our recent results indicate that U1 snRNP is indeed very important to the early stages of spliceosome assembly. Surprisingly, we found that it participates in the recognition of both splice sites, but that this recognition is not sufficient to define the specificity with which the splicing machinery cleaves the 5′ splice site.

Genetics of Behavior

The other aspect of our work is concerned with circadian rhythms. The choice of *Drosophila* was dictated by such factors as its suitability for genetic analysis and the large amount of behavioral work done on this animal. For example, it was known that its circadian rhythms, as in all other systems examined to date, manifest temperature compensation: i.e., the circadian rhythm periods (approximately 24 hours) change little if at all with temperature.

The *period* (*per*) gene has a profound influence on the animal's circadian rhythms. Alleles of the gene speed up, slow down, or apparently eliminate the animal's circadian rhythms. At least two different rhythms, including those of locomotor activity (the animal's analogue of our sleep-wake cycle), are similarly affected. This work suggested that the *per* gene might constitute an entrée into the mysterious world of the circadian oscillator.

We have focused on the *per* gene and its product with the thought that it might code for a "clock molecule." This possibility is even stronger due to the characterization of a variety of mutant and transformed strains.

Our experiments have reinforced the suggestion that the gene is making a physiological rather than developmental contribution to the oscillator—i.e., it is carrying out a biochemical func-

tion necessary to the running of the oscillator rather than contributing to the "building" of the oscillator during development. Until the *per* gene is turned on, the animal cannot tell time. These experiments also provide some indication as to where the circadian oscillator may reside.

We have also discovered that the *per* gene product cycles in a circadian fashion: it is present at relatively high levels in the middle of the day. Moreover, the *per* mRNA exhibits a similar "behavior"; the cycling of the *per* protein is due, at least in part, to a cycling of its mRNA. A variety of experimental approaches indicate that the *per* protein feeds back to influence the cycling of its own mRNA. Recent evidence suggests that a second post-transcriptional regulatory event is also necessary for cycling of the *per* protein. We suspect that this important feedback loop underlies the behavioral rhythms that are observed in normal and mutant fruit flies.

Determinants of Developmental Programs of Gene Activation

Michael G. Rosenfeld, M.D.—Investigator

Dr. Rosenfeld is also Professor of Medicine in the Eukaryotic Regulatory Biology Program at the School of Medicine of the University of California, San Diego. 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 the NIH. Dr. Rosenfeld also holds an adjunct position at the Salk Institute.

OVER the past year the central research focus in this laboratory has been to define the molecular mechanisms that dictate the developmental and regulated expression of neuroendocrine genes and to begin to apply emerging principles to analysis of neuronal 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 eukaryotic 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 specific genes encoding regulatory genes, receptors, and cell-specific transcription factors have permitted an initial definition of developmental and regulatory transcriptional and post-transcriptional strategies in the neuroendocrine system.

We are using the anterior pituitary gland as a model to investigate the molecular basis for generation of specific cell phenotypes in an organ. The rat genes for growth hormone and prolactin (a pituitary hormone that stimulates and sustains lactation) exhibit precisely restricted expression in somatotrophs and lactotrophs, respectively. Prolactin gene expression is dictated by two separate regions, a distal enhancer and a proximal region, each containing at least four critical elements. The two regions, each capable of targeting tissue-specific gene expression, act synergistically to generate high levels of prolactin gene expression in transgenic mice.

The consensus binding site for cell-specific transcription-regulating protein(s), referred to as Pit-1, was characterized; the 33-kDa Pit-1 protein was purified; and, using *in vitro* transcription assays, it was shown to exert functional effects. Pit-1 cDNAs were cloned from rat pituitary expression libraries. Bacterially expressed Pit-1 bound, specifically and with high affinity, 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.

The carboxyl terminus of Pit-1 contains a 60–amino acid region similar to the homeodomain of several *Drosophila* and vertebrate regulatory genes. Pit-1 contains a second, 76–amino acid sequence with striking homology to other transcription regulators. These two regions, together with a nonconserved region between them, are referred to as the POU domain. 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. The major transcription-activating domain of Pit-1 is a 70–amino acid amino-terminal region rich in serine and threonine and distinct from recognized motifs. A detailed structure-function analysis has suggested that the determinants of DNA binding by POU-domain proteins are distinct from those of the classical homeodomain proteins. 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 appears that the POU-domain protein acts as a developmental regulator to determine patterns of commitment, progression, and proliferation of specific cell types in the anterior pituitary gland.

We devised a strategy to isolate new members of the POU-domain gene family. Four were identified from cDNA in brain and testes. Transcripts from the *Brn-1* and *Brn-2* genes exhibited virtually identical patterns of expression in the central nervous system, although *Brn-1* was clearly expressed in the medullary zone of the kidney, and *Brn-2* was not. *Brn-3* mRNA was present predominantly in the peripheral nervous system; *Tst-1* transcripts were present in testes and brain. Subsequently, other POU domains have been identified expressed in early development or 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 was evident for all four transcripts at all levels. The time course of anatomic restriction in the developing neural tube was distinct, and the patterns for each gene product tended to reflect the adult loci of expression. In addition, reactive transcripts for some of these genes were also expressed during mammalian neurogenesis. We recently identified many additional family members in the brain and obtained evidence that at least some members of this family can bind to specific elements in distinct classes of neuronally expressed genes. Two POU-domain genes were cloned from *Drosophila* cDNA libraries and identified in neural structures during development. These proteins form stable heterodimers and act to alter patterns of activation of neuron-specific genes.

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.

We have provided evidence that a gene of the neuroendocrine system, the calcitonin/calcitonin gene-related peptide (CGRP) gene, contains genomic regions that represent discrete hormone-encoding domains. The ultimate expression of these domains depends on alternative RNA-processing events that differentially include or exclude specific exons encoding specific components of polypeptide regulators in the mature mRNA products. The rat and human calcitonin/CGRP genes are composed of 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. This appears to reflect the actions of a specific regulatory machinery controlling post-transcriptional RNA splice acceptor choice, dependent on a specific, discrete cis-active element.

In summary, these investigations have established several novel aspects of transcriptional and post-transcriptional regulatory strategies in neuroendocrine gene expression and in 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 approach involves studying genes whose mutations disrupt neural development. During the past year, we have focused 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*, which is an attractive system to study the mechanisms underlying lineage-independent developmental decisions, is a two-dimensional array of 800 repeating units, or ommatidia. Each ommatidium contains 8 photoreceptor cells and 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 this 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, cells that would normally become 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 express-

ing *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 determine how activation of the *sevenless* tyrosine kinase instructs 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 used two strategies in a genetic approach toward the identification of genes whose products act downstream of *sevenless*, including those that might be direct substrates for the *sevenless* kinase.

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 *sina* 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 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. 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.

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 patients treat transfused vWF as a foreign protein and make inhibitory antibodies to it. Deletions in the vWF gene appear to 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. Severe bleeding in certain variants can be associated either with increased or decreased binding of vWF to a specific receptor protein on platelets. Among nine such unrelated patients, five different mutations were characterized in a small vWF domain that appears to modulate the affinity of vWF for this platelet receptor. This illustrates the importance of proper balance between increased and decreased vWF function for normal hemostasis.

A newly described variant of von Willebrand disease, first recognized in 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. A missense mutation was identified in the factor VIII binding site of vWF. The corresponding recombinant vWF exhibited the same defect in factor VIII binding as natural vWF Normandy, confirming that this mutation causes the disease. The genetic defects of these and other such patients provide insights 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 to understand the abnormal blood coagulation that accompanies much human dis-

ease. We have cloned and expressed recombinant human thrombomodulin in a variety of cultured cell lines, and the recombinant protein appears to be structurally and functionally normal. These and similar cells that express altered forms of thrombomodulin are being used to investigate 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 anti- 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 pro- 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 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 regulated and will extend this work to other endothelial cell proteins that can promote or inhibit thrombosis.

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 higher organisms, a fertilized egg undergoes many divisions to produce a multicellular body with characteristic segments. The basic pattern of the body is invariant from generation to generation and is dictated by a genetic blueprint within the organism's own genome. My group is interested in learning about the genetic blueprint of the fruit fly *Drosophila melanogaster*.

The fruit fly's body is made up of several fused segments in the head, three thoracic segments with wings and legs, and ten abdominal segments. The unique characteristics of each segment are determined by the activities of the genes called homeotic. When these genes malfunction, a body segment or group of segments transforms to take on characteristics of another segment. For example, some mutations in the homeotic gene *Ultrabithorax* cause the transformation of a non-winged segment into one with wings, leading to a four-winged instead of the normal two-winged fly. Thus homeotic genes can be thought of as master regulators that trigger the genetic circuits necessary for the normal body pattern. Our goal is to understand how the homeotic genes accomplish this task at the molecular level.

Three homeotic genes—*Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*)—are responsible for determining the characteristics of two thoracic and eight abdominal segments. They are located in a chromosomal region called the bithorax complex (BX-C). Flies carrying mutations in any one of these BX-C genes clearly show characteristic transformations. By noting which segments are transformed in these mutants, we can infer that *Ubx* is active 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 about what the homeotic gene products do, we have focused our studies on *abd-A* and *Abd-B*. Since each of the abdominal segments shows unique characteristics, it seemed possible that these genes encode multiple protein products that are each responsible for determining the identity of a particular segment. Contrary to this expectation, we have found that *abd-A* and *Abd-B* encode one and two proteins, respectively. Each of these proteins contains a stretch of almost identical amino acid sequence, known as the homeodomain. This sequence is found in the proteins from many organisms, including humans.

As has been shown for other homeodomain-containing proteins, we have demonstrated that *abd-A* and *Abd-B* proteins bind to DNA. Furthermore, we have shown that the *abd-A* protein can repress the transcription of other genes in cultured *Drosophila* cells. It seems, therefore, that the *abd-A* and *Abd-B* proteins accomplish their tasks by regulating the transcription of other genes, which in turn contribute to unique characteristics of each segment.

We are now extending our work to demonstrate that the homeotic proteins regulate the transcription process in developing embryos as well as in tissue culture cells. For this purpose we are studying the interaction of *abd-A* with a transcription start site, or promoter, for another homeotic gene, *Antennapedia* (*Antp*). Genetic evidence suggests that *Antp* expression in wild-type embryos is suppressed in the first abdominal segment by *Ubx* and in the second-through-seventh abdominal segments by *abd-A*. From the genetic evidence, however, it is not known if this suppression is direct or indirect. We have now shown that purified *Ubx* and *abd-A* proteins bind to a number of sites near the promoter of the *Antp* gene.

To test if these *in vitro* binding sites are indeed utilized for repression of *Antp* in embryos, we have mutagenized each binding site, rendering it incapable of binding the homeotic proteins. These altered DNA sequences are being introduced into the genome of the fly to see if the normally observed repression of *Antp* by *Ubx* or *abd-A* protein is abolished. If so, this would provide the first conclusive evidence that homeotic proteins act as transcriptional regulators in developing embryos.

We are also interested in knowing 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, leading to different combinations of homeotic proteins in the various 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 present in the fifth-through-eighth abdominal segments. These three combinations would of course define only three segmental 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, for example, that some cells within the fifth abdominal segment express *abd-A* protein exclusively or predominantly; other cells express *Ubx* protein; and yet other cells, *Abd-B* protein. In the sixth abdominal segment, *Ubx* and *abd-A* expression appears similar to that of the fifth abdominal segment, but *Abd-B* expression is different from that of the adjacent segment. This kind of "mosaic" expression can, in theory, specify an unlimited number of segmental identities. We have therefore asked the question, Is the mosaic expression necessary for correct specification of segment identities?

To answer this, we constructed genetically altered flies in which we could express *Abd-B* protein at will. When *Abd-B* protein is expressed in all cells, abolishing the normal mosaic pattern of *Abd-B* expression, thoracic and abdominal segments exhibit characteristics of the eighth abdominal segment. Surprisingly, we do not observe completely nonsensical identities, as might have been expected from this abnormal situation. One reason for this result could be that the eighth segment identity requires only *Abd-B* protein (although *Ubx* and *abd-A* proteins are present at low levels) and that the high level of *Abd-B* protein in other parts of the body is sufficient to suppress the effects of *Ubx* and *abd-A*, leading to a repetition of the eighth abdominal segment identities. This suppression of *Ubx* and *abd-A* activities could be due to repression of *Ubx* and *abd-A* transcription by *Abd-B*, or alternatively, the *Abd-B* protein could simply override the effect of *Ubx* or *abd-A* protein in the same cell.

To distinguish these possibilities, we asked what segmental identity would result if we engineered the fly so that both *Ubx* and *Abd-B* proteins are expressed in all cells. Our preliminary results indicate that *Ubx* protein exerts its effect over *Abd-B* protein in anterior body segments, while in posterior segments the reverse applies: the *Abd-B* protein predominates. Since the level of *Ubx* or *Abd-B* proteins should be the same in all cells, these results suggest that the response depends on where the cell is located. This in turn suggests that there may be other factors in addition to the homeotic genes that contribute to the correct identities of body segments.

Intracellular Transport of Proteins

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.

RESEARCH in our laboratory is devoted to a molecular description of two processes: polypeptide translocation from the cytosol into the endoplasmic reticulum (ER) and vesicular transport among organelles of the secretory pathway.

Genetic Dissection of the Secretory Process

A genetic approach to the study of protein transport in eukaryotes 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. Over 30 gene products have been implicated in the process of delivering membrane and secretory proteins to the cell surface.

A combination of 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.

Protein Translocation Into the Endoplasmic Reticulum

Protein translocation into the lumen of the ER 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.

Four *SEC* genes have been identified whose gene products are required for the translocation of a wide variety of secretory and membrane proteins. Three of the Sec proteins (Sec61p, -62p, and -63p) are found in a complex that may be isolated from detergent-solubilized membranes. The complex includes two other polypeptides of unknown structure. This complex may catalyze the translocation event or create a pore in the ER membrane through which secretory polypeptides are transmitted.

At least one of the proteins in the translocation complex forms a direct contact with the nascent secretory polypeptide. A protein conjugate consisting of the yeast secretory protein, α-factor precursor, linked chemically to avidin, engages but jams the translocation apparatus. Treatment of the jammed complex with a cleavable cross-linking reagent generates a covalent connection between α-factor precursor and Sec61p. The requirements for the production of this intermediate suggest that the contact between Sec61p and the secretory polypeptide follows an earlier event in which α-factor first binds to a receptor on the cytosolic surface of the ER membrane.

A fourth gene, *SEC65*, required for translocation, resembles the 19-kDa subunit of the mammalian signal recognition particle (SRP). *sec65* mutants are defective in the translocation of molecules that are inserted cotranslationally, such as yeast invertase, or post-translationally, such as α-factor precursor. This implies that at least part of a yeast SRP participates in the translocation of molecules uncoupled from ongoing protein synthesis.

Two forms of the 70-kDa heat-shock protein family (heat-shock cognate, or Hsc70) also participate in the translocation event. Depletion of cytosolic Hsc70 by genetic manipulation results in the accumulation of certain secretory and mitochondrial precursor proteins. A direct participation of these molecules in the translocation event was established by reconstitution of Hsc70-dependent *in vitro* α-factor precursor assembly into yeast ER vesicles. Cytosolic Hsc70 may associate with translocation precursor polypeptides and prevent them from folding into unfavorable structures or complexes.

A luminal Hsc70, homologous to mammalian BiP, was shown by Joe Vogel and Mark Rose at Princeton also to be involved in translocation.

Luminal Hsc70 may exert its influence through interaction with the integral membrane complex of Sec proteins. Sec63p, which is a member of this complex, has a luminal domain with sequence homology to a bacterial heat-shock protein known to interact with the *Escherichia coli* homologue of Hsc70. Luminal Hsc70 may control the assembly and disassembly of the translocation complex by binding to Sec63p.

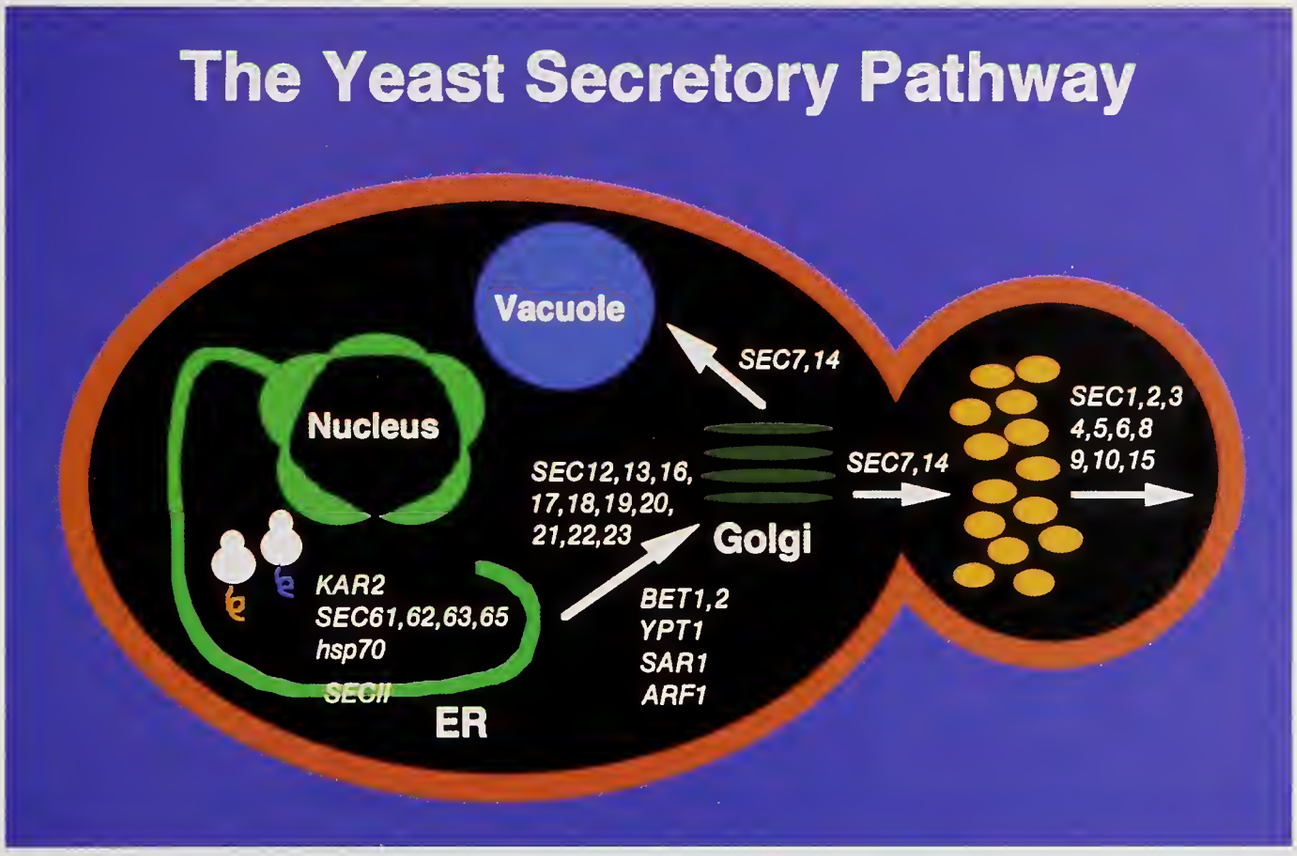
Vesicle Transport Early in the Secretory Pathway

Subsequent stages in the secretory pathway involve protein sorting and transport from the ER 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. There are two categories of *sec* mutants defective in this limb of the pathway: mutant cells that accumulate ER tubules at a restrictive temperature (class I; *sec12*, *-13*, *-16*, and *-23*) and mutant cells that also accumulate several thousand 60-nm vesicles (class II; *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 gene 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 genes and of class 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.

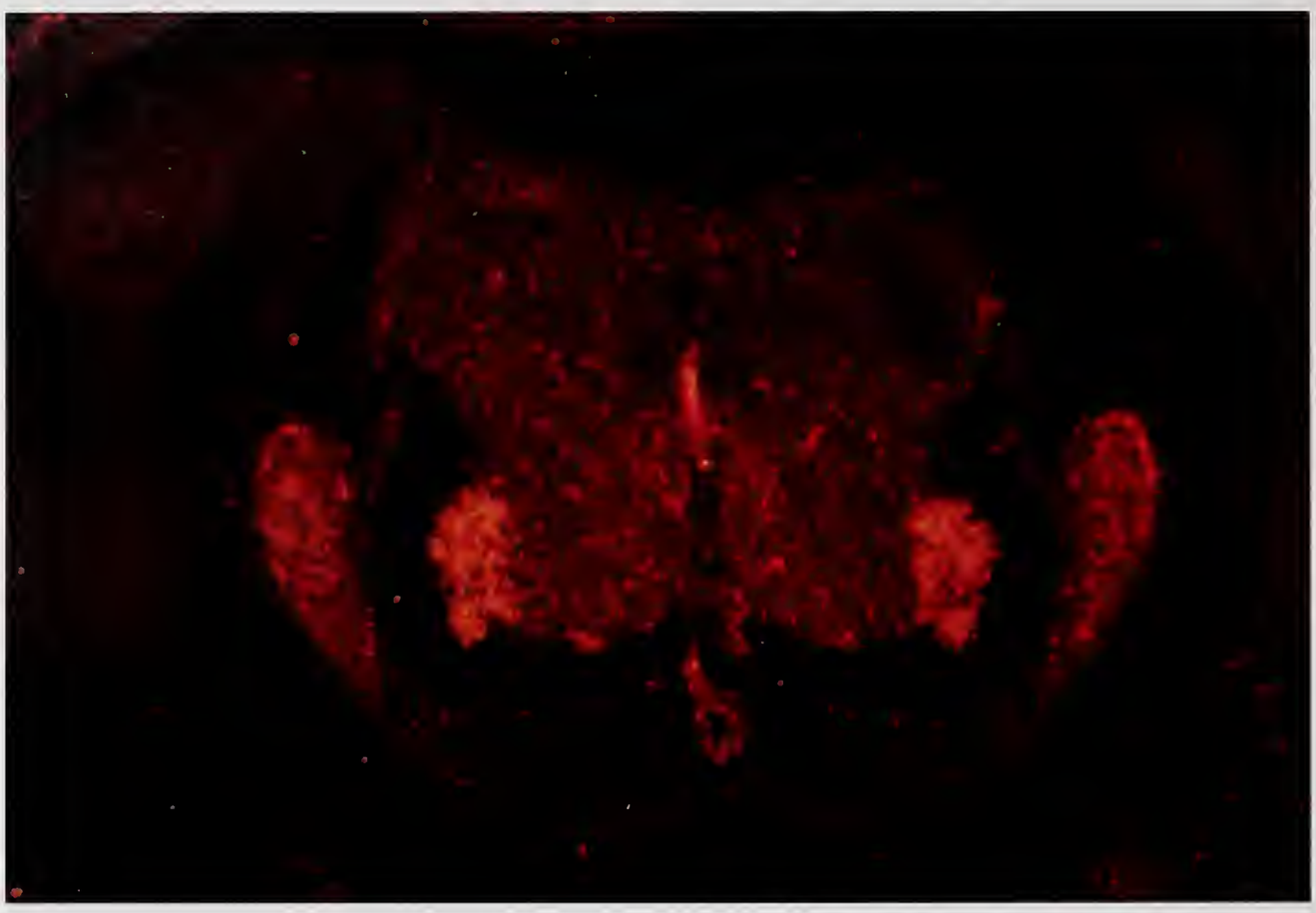
Transport of α-factor precursor *in vitro* is mediated by diffusible vesicles. Transport vesicles contain core-glycosylated precursor and are physically separable from donor ER and target Golgi membranes. Budding of vesicles from the ER requires a crude cytosol fraction, ATP, Sec12p, Sec23p, and GTP. Fusion of the vesicles to the Golgi compartment is measured by the conversion of core-glycosylated 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. Additional proteins in the cytosol fraction that facilitate vesicle budding are being purified.

Vesicle budding in the secretory pathway was thought to be mediated by clathrin, a structure-forming protein that spontaneously assembles into coats about the size of small transport vesicles. A test of clathrin function was performed by molecular cloning and disruption of the chromosomal genes encoding the large and small subunits of the protein. Clathrin is important but not essential for cell growth. Deletion mutations that eliminate clathrin produce sickly cells that nevertheless are competent for most aspects of protein transport. A specific lesion in α-factor precursor processing, an event that occurs in the Golgi apparatus, is deficient in clathrin mutant cells. Alternative structure-forming proteins that serve more essential roles in secretion may be discovered using the Sec protein–dependent vesicle-budding reaction.



SEC *(secretory) genes regulate the flow of yeast secretory polypeptides into and from the endoplasmic reticulum (ER) and through the Golgi complex. In the Golgi, proteins destined for secretion are sorted from vacuolar/lysosomal proteins. Mature secretory proteins are packaged into vesicles and directed to the inner surface of the growing yeast bud. The secretory process and the* SEC *gene products are very similar in yeast and mammalian cells.*

Research of Randy W. Schekman.



Analysis of embryonic rat motor neurons demonstrates that these cells express a gene encoding agrin. The agrin molecule is secreted by the nerve cell axons as they contact the muscle fibers. The agrin protein plays a role in organizing the synapse and may be important in regeneration following nerve or muscle damage.

From Rupp, F., Payan, D.G., Magill-Solc, C., Cowan, D., and Scheller, R.H. 1991. Neuron *6:811–823.*

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. Dr. Scheller received his B.S. degree from the University of Wisconsin–Madison, and his Ph.D. degree in chemistry from the California Institute of Technology. He 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

Most synapses release two types of chemical messengers: fast-acting compounds, 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, while the other region of the precursor is not stored for secretion. Thus the AtT-20 cells, like the bag cells, differentially route the two regions of the ELH prohormone. As a result, we have a cell culture system for investigating the mechanism of differential routing through the secretory pathway. We have generated mutations in the first processing site, which is thought to be critical for the sorting. The effects of these mutations are currently being analyzed.

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.

To study the functions of these proteins, we are using DNA transfection and antibody techniques. Wild-type and mutated forms of the proteins are being transfected into AtT-20 and PC12 cells. In addition, some of the proteins have homologues in *Drosophila*. We have constructed vectors containing wild-type or mutated forms of the proteins driven by heat-shock promoters. These constructs allow us to transfer the gene into flies and to turn on expression at any time during development and in the adult.

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. We are currently attempting to express this gene in cell culture to study the roles of the agrin molecule in the formation of synapses.

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 post-doctoral fellow with Emil Gotschlich at the Rockefeller University. He founded the Division of Geographic Medicine at Stanford University and 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 my laboratory is to discover how these infectious agents cause disease, how they are spread between individuals, and how this information can lead to new tactics for prevention and treatment. 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 the organism attaches to, invades, and damages human cells. 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, where its reservoirs are, and how it manages to survive as a viable entity in the real world.

From studies of the latter kind, we have learned that the survival of a pathogenic microorganism correlates with its ability to occupy different 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. From an experimental point of view, we have tried in the last year to identify molecular and genetic events that transpire within an infectious agent as it enters a new habitat and adapts.

The results of these studies have led to the recognition that pathogenic bacteria seldom attack the gastrointestinal tract as individual, independent, well-separated organisms. Instead, they first seem to form infectious units composed of multiple organisms and then become established as adherent colonies on the surface of the intestine. While our studies have focused on one kind of microbe—enteropathogenic strains of *Escherichia coli* (EPEC)—we have discovered that this phenomenon is exhibited by many other bacterial species that selectively infect mucosal surfaces.

Accordingly, we now believe that the capacity to form infectious units confers a selective advantage on bacterial species infecting mucous membranes, that it is very probably an essential determinant of virulence, and that it offers a logical point of attack for the prevention of disease by vaccination. This phenomenon is discussed below as it pertains to EPEC and its ability to cause gastroenteritis.

Formation of Infectious Units

EPEC are a common cause of infantile diarrhea in Third World children. When small bowel biopsies of infected children are performed, colonies of EPEC are found attached to the underlying epithelia. Similar findings are also seen in cultured epithelial cells, to which EPEC readily adhere as circumscribed clusters of bacteria, a phenomenon known as localized adherence. From these observations we know that the bacteria within the colonies not only interact with host cells to which they are bound, but also with each other. Scanning electron micrographs revealed multiple, ribbon-like structures coursing between the bacteria, and fibers also appeared to tether individual bacteria to the epithelial cell surface.

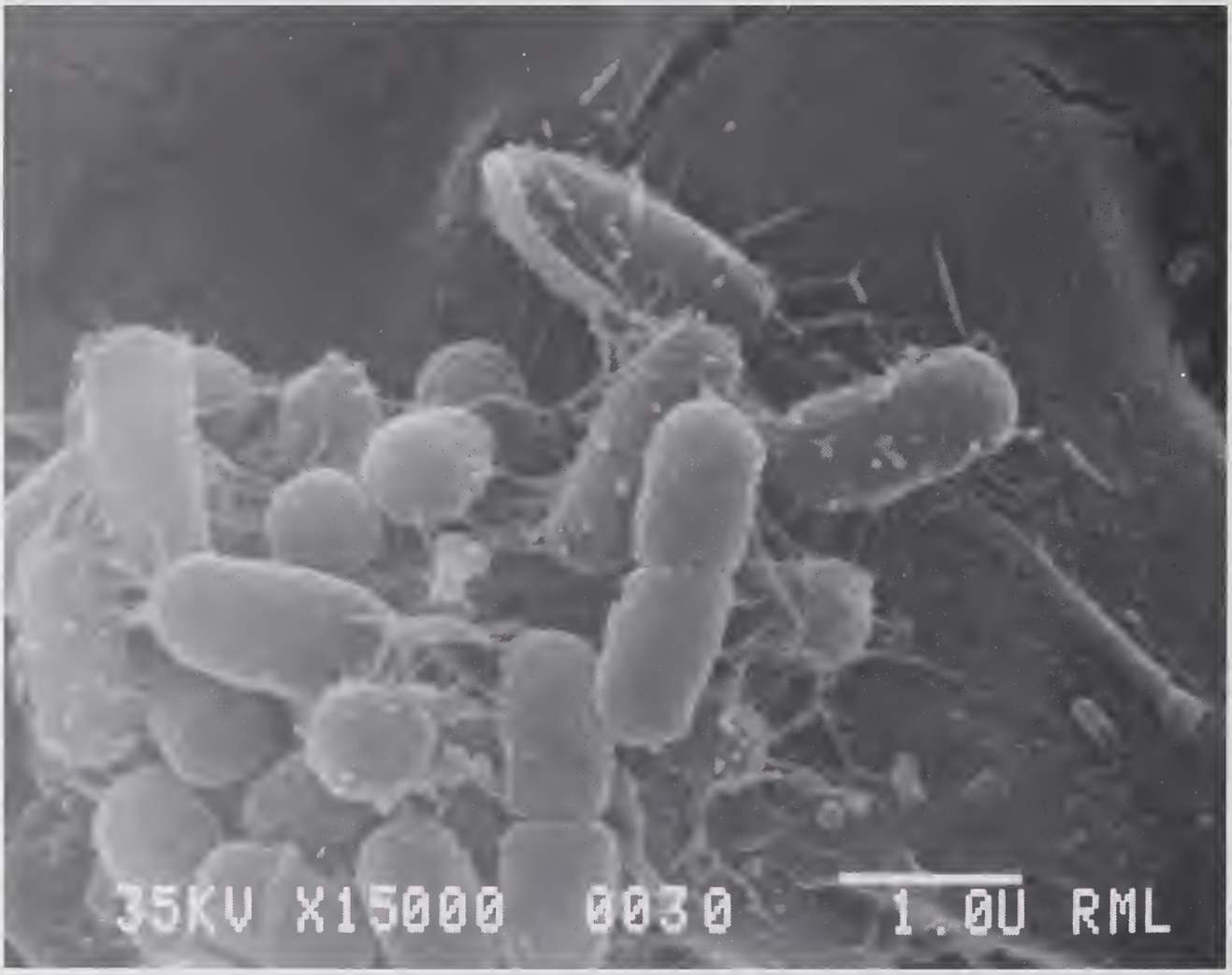
Transmission electron micrographs of these interbacterial ribbons showed that they are comprised of many individual filaments that, having laterally aggregated and intertwined, formed bundles 50–500 nm wide and 15–20 μm long. These tended to twist, curl, and form loops. Moreover, bundles expressed by different organisms appeared to form three-dimensional arrays resembling a fishing net or a meshwork, in which individual bacteria were embedded.

EPEC did not express these bundles while growing on most laboratory culture media. However, when transferred to chambers containing cultured epithelial cells, EPEC evinced bundles within 15 minutes, and by 30 minutes colonies composed of tightly adherent bacteria appeared

on the epithelial cell surfaces as on the intestinal tissue of an infected child. Thus the expression of bundles by EPEC is induced by physical and chemical variables found near the epithelial cell surface, where the bacteria seem to colonize through their propensity to form a network linking one to another. We have elected to call these structures colony-forming bundles, or CFB. It now seems likely that the colony begins to form as the bacteria descend to the epithelial cell surface and that it therefore constitutes the EPEC infectious unit.

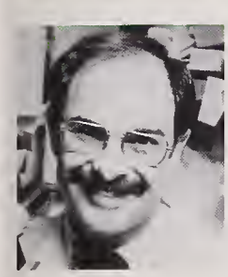
The role hypothesized for CFB was corroborated by two additional observations. First, deletion of genes from EPEC that are required for the production of CFB yielded mutant bacteria that were unable to form cell surface colonies. Second, antibodies to CFB substantially reduced the colony-forming capacity of EPEC strains. This latter finding suggests the use of a CFB vaccine to prevent diarrhea.

We determined the molecular structure of the CFB filaments. Remarkable similarities were found between their amino acid sequence and the sequences of fibrous appendages produced by a variety of pathogenic bacteria that infect mucous membranes, including the agents of cholera and gonorrhea. Thus the CFB appears to be a shared structural innovation by bacterial species that live on mucous membranes.



Scanning electron micrograph of enteropathogenic Escherichia coli *(EPEC) growing on an epithelial cell and expressing colony-forming bundles.*
Research of Gary K. Schoolnik.

Lymphocyte Surface Antigens in Health and Disease



Benjamin D. Schwartz, M.D., Ph.D.—Investigator

Dr. Schwartz is also Professor of Medicine and of Molecular Microbiology at Washington University School of Medicine, St. Louis, and Chief of the Division of Rheumatology at Jewish Hospital, Washington University Medical Center. He received his B.A. degree in mathematics from Columbia College and his M.D. and Ph.D. degrees in immunology from Albert Einstein College of Medicine. Specializing in immunology, he was a postdoctoral research fellow with William Paul in the Laboratory of Immunology at the NIH. Dr. Schwartz is a member of the American Federation for Clinical Research, the American Society for Clinical Investigation, and the Association of American Physicians, and served as President of the American Federation for Clinical Research.

IN the initiation of an immune response against a foreign protein antigen, the antigen is taken up by a specialized antigen-presenting cell (APC) and broken down into small peptide fragments. These fragments are then thought to be bound by proteins of the immune system, namely major histocompatibility class II molecules. The complex of class II molecule and bound peptide fragment is presumably expressed on the APC surface, where it can be recognized by a specific T lymphocyte. The recognition event activates the T lymphocyte, which then initiates the immune response.

Despite a large body of circumstantial evidence from which the presence of class II molecule linked to a foreign peptide on the APC has been inferred, the existence of this complex on the surface of viable cells has not been directly demonstrated.

We have been collaborating with Tom and Vivian Braciale in the Department of Pathology at Washington University School of Medicine to elucidate the immune response against influenza virus. A particular human T cell clone, termed V1, was demonstrated to recognize a synthetic peptide, HA 128–145 (corresponding to an 18–amino acid segment of the influenza viral hemagglutinin protein), only when that peptide is presented on an APC bearing a particular class II molecule, HLA-DRw11.

The recognition of this complex by the T cell clone indicated that the HA 128–145 peptide was immunogenic. We therefore used the HA to immunize a rabbit and raise an antiserum, anti-HA. The antiserum recognizes not only free HA peptide but also a complex of the peptide bound to HLA-DRw11 on the surface of an APC.

The composition of this complex was proved in several ways. Anti-HA did not recognize a mutant APC that expressed no class II molecules after the cells were exposed to HA, demonstrating that the latter did not bind nonspecifically to APCs. Also, anti-HA did not bind to DRw11-bearing APCs that had not been exposed to HA, nor to APCs bearing several DRs that had been exposed to the peptide.

A second APC, known to express only half the amount of DRw11 as did the first APC bearing DRw11, when exposed to HA bound only half the amount of anti-HA. This last result indicates that binding of anti-HA to complexed DRw11-HAs is commensurate with the level of DRw11 expression.

Further proof of the composition of the complex was achieved biochemically. The HA peptide was tagged with radioactive iodine and a photoreactive group, so that when exposed to ultraviolet light, the photoreactive group on the peptide would form a covalent bond with any molecule to which the peptide had bound. DRw11-bearing APCs were incubated with the modified peptide, exposed to ultraviolet light, and then lysed. The cell lysate was then reacted with either the anti-HA antibody or a monoclonal antibody that recognized the DRw11 molecules.

Biochemical analysis of the molecules reactive with the antibodies demonstrated by several criteria that the molecules covalently bound to the radioactive iodinated peptide were indeed the DRw11 molecules. Furthermore, the DRw11 were the only molecules tagged with radioactive iodine, demonstrating that no other molecules of the APC were capable of binding peptide.

A series of experiments was done with metabolic inhibitors to ascertain whether the HA 128–145 peptide followed the proposed route of intracellular processing. These experiments demonstrated that metabolic inhibition, including poisoning and fixation of the APC, did not decrease the number of DRw11-HA complexes detected by anti-HA. These results therefore indicate that HA must be binding directly to cell surface DRw11 molecules. It is of interest, however, that a temperature of 4°C totally abrogated, and 18°C significantly inhibited, the HA peptide's binding to DRw11 molecules. Collectively these

results suggest that temperature directly affects the ability of DRw11 molecules to bind HA 128–145.

Because these results suggested that the peptide was not following the normal antigen-processing pathways, we turned to a system where we knew these pathways were being used. It had already been shown that the T cell clone V1 could recognize naturally processed antigenic peptide derived from an intact influenza virion on the surface of a DRw11-bearing APC when the APC was exposed to intact virions. We therefore tested whether anti-HA could inhibit this recognition.

Anti-HA indeed inhibited recognition by 80 percent. This result established that anti-HA could recognize the peptide naturally derived from the intact virion. In addition, we have recently synthesized a longer peptide, encompassing 45 amino acids, including HA 128–145 (which we believe has to be processed by a DRw11-bearing APC to be recognized), and have demonstrated that anti-HA will bind this complex.

These results indicate that anti-HA can be utilized to identify naturally processed hemagglutinin peptide, as well as the compartments within the APC where processing of intact antigen to peptide occurs. These studies are ongoing.

Knowledge of such antigen-processing pathways will allow us to manipulate an immune response so that it may be augmented during vaccinations and suppressed in immune responses against self that produce autoimmune disease.

Second Messenger Pathways in Identified Neurons

James H. Schwartz, M.D., Ph.D.—Investigator

Dr. Schwartz is also Professor of Physiology and Cellular Biophysics and of Neurology at Columbia University College of Physicians and Surgeons. He received his M.D. degree from New York University Medical Center and his Ph.D. degree from the Rockefeller University for work with Fritz Lipmann on the catalytic site of alkaline phosphatase in Escherichia coli. *He then joined the faculty of the Department of Microbiology at New York University. While there, he joined the Division of Neurobiology and Behavior, which later moved to Columbia's College of Physicians and Surgeons. Dr. Schwartz's honors include the Selman A. Waksman Award in Microbiology. He has been a Visiting Scholar at the New York Psychoanalytic Institute and at the NIH.*

STIMULATION of a nerve cell can result in long-lasting changes in its biochemistry. These are produced by a process called signal transduction. The stimulus, delivered by either a neurotransmitter or a hormone, acts upon specific receptors on the nerve cell's surface; and these, in turn, induce the formation of "second messengers" inside the cell.

Second messengers usually work by binding to proteins that might be called secondary effectors. In many instances the effectors are enzymes termed protein kinases. When activated, protein kinases catalyze the transfer of a phosphoryl group from ATP to a large number of different cellular proteins. As a consequence of the phosphorylation reactions, the functions of these cellular proteins are changed. Thus production of the second messenger alters the way the neuron behaves.

How long the nerve cell remains changed initially depends on the lifetime of the second messenger. The secondary effector remains activated as long as there is enough second messenger within the cell. After brief stimulation, second messengers are formed for only a few minutes, the action of the secondary effector is brief, and the altered state of the neuron persists for less than an hour. If the stimulation is prolonged, however, the result is quite different. An interesting and important phenomenon occurs.

In such cases, the activity of the secondary effector enzyme persists even after formation of the second messenger stops. One might say fancifully that the enzyme is taught to behave differently. Like the animal, the molecule behaves as if it has learned. This phenomenon underlies the enhancement of synaptic transmission believed to be the basis of several simple kinds of learning and memory.

We have been interested in the mechanisms by which secondary effector enzymes become educated. In previous years, we reported progress in understanding the mechanisms for three of the best known second messenger pathways. After discovering in 1985 the phenomenon for the Ca^{2+}/calmodulin-dependent protein kinase (a secondary effector enzyme for Ca^{2+}), we described similar phenomena for the cAMP-dependent protein kinase (PKA) and for protein kinase C. This year we made significant advances in understanding how PKA is made to persist for many hours after stimulation ceases.

For all of these protein kinases (which are distantly related in evolution), the second messenger activates the enzyme by releasing an inhibition built into the latter molecule. This process is reversible. With PKA the inactive form is a tetrameric complex consisting of two identical regulatory (R) subunits that fit precisely together with two catalytic (C) subunits. The R subunits, upon binding the second messenger, cAMP, release the two C subunits for action.

In earlier experiments we demonstrated that the ratio of R to C changes in sensory neurons of the marine mollusk *Aplysia californica* to cause the enhanced protein phosphorylation that occurs in the presynaptic facilitation of sensory-to-motor neuron synapses in long-term sensitization of defensive reflexes. The change consists of a decrease in the amount of R subunit, which requires the synthesis of new protein.

What is the mechanism for producing the change in the R to C ratio observed in sensory neurons of *Aplysia* trained behaviorally to exhibit the long-term sensitization? To answer this question, we cloned cDNAs encoding both R and C subunits and therefore can determine the amounts of their messenger RNAs. In addition, we can raise antibodies to measure the amounts of subunit protein in sensory cells before, during, and after training. We find that the PKA changes after long-term training because R subunits in the neurons disappear. Furthermore, there is no change in the amount of C subunits, nor in R or C subunit messenger RNAs.

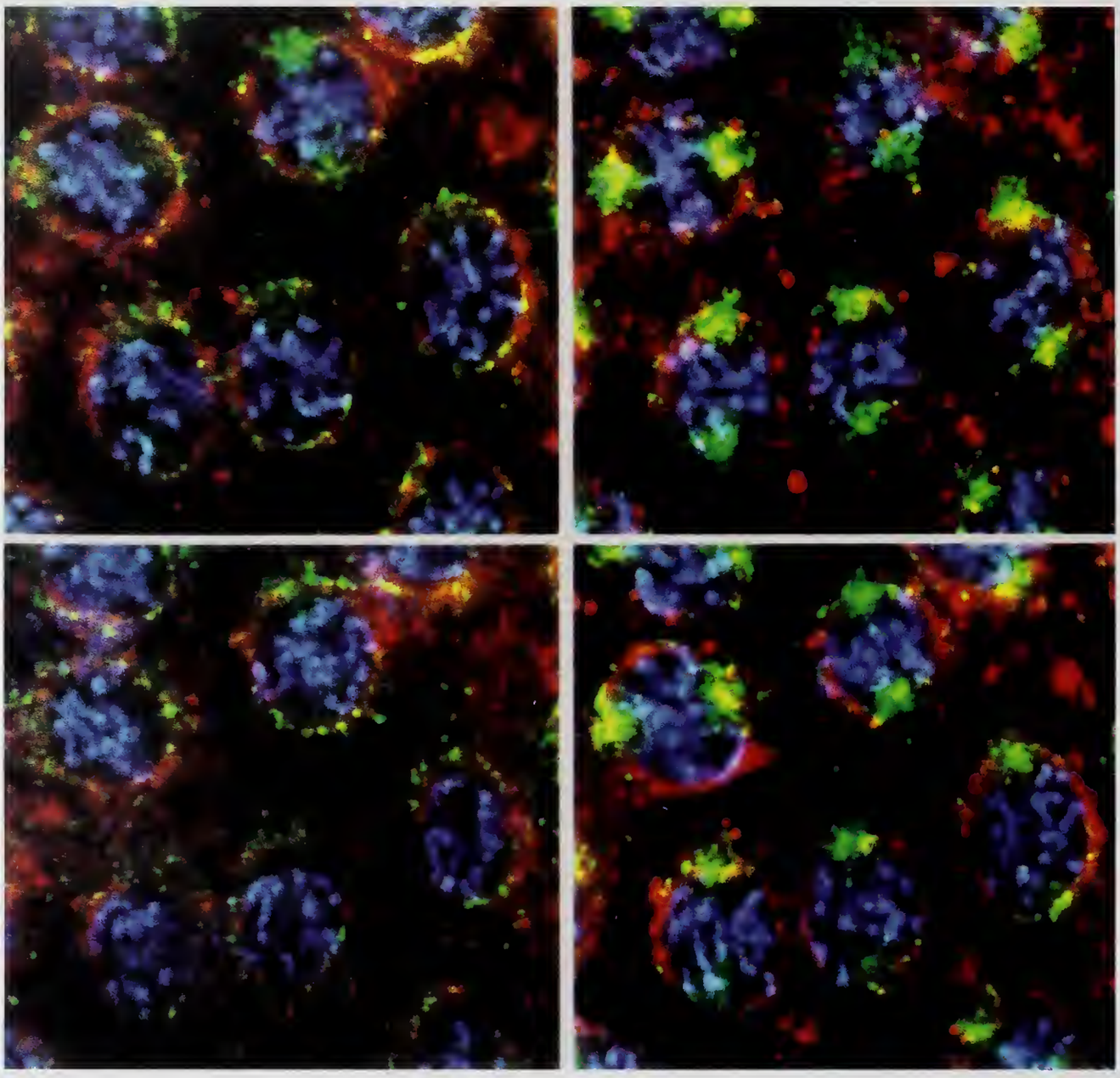
Disappearance of proteins is commonly caused by enzymes that digest them. In the sequence of amino acids of the R subunits inferred from the cDNA cloning experiments, there is an amino-

terminal domain that is highly enriched in proline, glutamine, serine, and threonine (PEST) residues. This sort of domain has been recognized as a determinant of proteins that turn over rapidly. But digestion of proteins does not normally require the synthesis of new protein. Why then does the degradation of R subunits depend on new protein synthesis?

Preliminary evidence suggests that the proteolytic mechanism responsible for the disappearance of R subunits under conditions of training depends upon ubiquitination, a complex degradative process involving enzyme-catalyzed coupling of a small protein, ubiquitin, to the protein to be destroyed.

Since ubiquitin and the enzymes required for coupling are present in the neuron whether or not it has been stimulated for long-term facilitation, how is the degradation of R subunits enhanced in stimulated neurons? Our working hypothesis is that the new protein synthesized as a consequence of long-term stimulation is a factor that targets R subunits for coupling with ubiquitin.

A similar hypothesis has recently been advanced for the periodic degradation of cyclin, a protein that regulates a special protein kinase crucially involved in cell division. In early frog embryos, cyclin is rapidly proteolyzed every 30 minutes during mitosis. We find that R subunits of PKA in early *Aplysia* embryos also are degraded through ubiquitin-mediated proteolysis. In contrast, the R subunit of mature muscle is quite stable. These observations that R subunits are readily down-regulated in growing cells and in nervous tissue provide support for the popular idea that both learning and development employ similar molecular mechanisms.



Superimposed images of chromosomes, microtubules, and the nuclear envelope. Optical sections of embryos triply stained for these entities were obtained at 0.2-μm focus intervals. Selected optical sections are displayed, each separated by 1 μm along the focus direction from the interior (lower left) to the surface (upper right) of the embryo. Chromosomes, microtubules, and nuclear envelope are displayed in blue, green, and red, respectively.

Research of John W. Sedat, David A. Agard, and Yashui Hiraoka.

Three-Dimensional Structure of Eukaryotic Chromosomes

John W. Sedat, Ph.D.—Investigator

Dr. Sedat is also Professor in the Department 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 David Agard (HHMI, University of California, San Francisco), has investigated chromosome structure from the perspective of a series of interlocking questions. 1) What are the levels of chromosome architecture 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 chromosome 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.

Although 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, we continue to perfect and enhance the instrumentation. We have improved the time resolution for data collection in the four-dimensional image work, 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 intricate three- or four-dimensional data.

Four-Dimensional Optical Microscopy

We have continued to study the structure of the cell nucleus in living *Drosophila* embryos. Nuclei were labeled by microinjection of fluorescent histones or other abundant chromosomal proteins. Nuclear and chromosome structures were followed throughout the cell cycle during embryonic development. In addition to discerning structural changes, we could now infer function.

This approach has shown several discrete chromosome sites attached to the nuclear envelope. These sites are the last to decondense during telophase (the stage in the nuclear division cycle that follows chromosome separation). Similar sites are observed at the beginning of prophase (the stage of nuclear division involving chromosome condensation), with condensation proceeding bidirectionally away from these centers.

During prophase, a remarkable wave of chromosome compaction started at the centromeres (chromosome regions at which mitotic spindles attach) and proceeded toward the opposite nucleus pole that contained the telomeres (ends of chromosomes). We could also determine that the site for the start of the compaction wave coincided temporally and spatially with the initial breakdown of the nuclear envelope as well as the site of the mitotic spindle formation, suggesting that there is a high degree of temporal-spatial control and organization in the mitotic process.

A major conclusion is that the structures in real-time living nuclei were very similar, if not identical, to the three-dimensional nuclear structures determined in our ongoing studies at higher resolution in fixed embryos.

Three-Dimensional *In Situ* Hybridization

We have analyzed the spatial arrangement of chromosomes in embryos of *Drosophila*. One specific biological question is whether homologous chromosomes (one from the male parent, the other from the female) have an ordered arrangement in a diploid nucleus. Our previous results showed that homologous chromosomes are not associated from prophase to anaphase in syncytial blastoderm stage embryos. This apparently contradicts genetic evidence for transvection, which suggests that the homologous chromosomes are associated. Thus we wished to test whether homologous chromosomes are asso-

ciated in interphase, when the essential processes of gene transcription and DNA replication occur.

We used *in situ* hybridization techniques to determine specific chromosomal regions within a nucleus. The location of the hybridization signals is revealed by labeling these embryos with a fluorescently tagged molecule. High-resolution three-dimensional optical sectioning microscopy of such embryos reveals the location of the nuclear DNA sequences.

We first analyzed the nuclear location of the histone gene cluster that lies close to the centromeric heterochromatin on chromosome 2L. Our results, using a histone gene probe, revealed that homologous chromosomes are also separated in the majority of interphase nuclei at syncytial blastoderm-stage embryos up to the 13th nuclear cycle. In dramatic contrast, at the 14th nuclear cycle when cellularization begins, homologous chromosomes are associated in the majority of nuclei. The frequency of homologue pairing reaches the maximum (about 95 percent) by the time of gastrulation.

Analysis of the three-dimensional location of histone hybridization signals showed that histone genes are located on the nuclear midline at the 13th nuclear cycle and move toward the centromeric cluster on the apical side of nuclei at the 14th nuclear cycle. This can perhaps be explained by the formation, at the 14th nuclear cycle, of heterochromatin, which process may be involved in initiating the pairing of homologous chromosomes. Translocations further define homologue pairing, with strong effects on homologous chromosome association.

We are currently using a wide range of DNA probes for specific chromosomal sequences to study processes of homologue association. Our preliminary results suggest that frequency and timing of association vary from one chromosomal locus to another. For example, the gene *engrailed* exhibits homologue association at the 14th nuclear cycle with high frequency, but the gene *Ubx* does not. A generalized picture will be obtained once a larger number of genetic loci are analyzed.

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 reforms during telophase?

We are localizing the other two well-known components of the nuclear periphery relative to the lamin network, using monoclonal antibodies directed against a major glycoprotein component of nuclear pore complexes (GP-190) and the DNA-specific stain DAPI. 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 directly contacting lamins. The majority of it seems to be at a distance of about 0.2 μm. This result is consistent with much indirect evidence of a strong interaction between chromatin and the nuclear lamina, but suggests strongly that a direct physical contact is not involved.

We are injecting 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. 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 functional assays will be required for proper interpretation of the biochemistry.

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.

Molecular Basis of Familial Hypertrophic Cardiomyopathy: Cardiac Myosin Heavy-Chain Gene Mutations

Jonathan G. Seidman, Ph.D.—Investigator

Dr. Seidman is also Professor of Genetics at Harvard Medical School. He received his undergraduate degree in biochemistry from Harvard University and his Ph.D. degree from the University of Wisconsin, where he studied with W. H. 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 myocardial hypertrophy, a wide spectrum of symptomatic involvement, and a 2–4 percent annual mortality rate from sudden death, which can occur without warning. Postmortem examination reveals increased myocardial mass with myocyte and myofibrillar disarray. Diagnosis is based on typical clinical features and the two-dimensional echocardiographic demonstration of unexplained left and/or right ventricular hypertrophy. In the young, diagnosis is often complicated, because hypertrophy may not develop until after adolescent growth has been completed. The anatomical distribution of myocardial hypertrophy and severity of symptoms may be quite variable, even within a family.

During the past 30 years, the cardiac features of FHC have been extensively reported, but the etiology and molecular pathophysiology have remained speculative. Last year we reported the localization of a gene on chromosome 14 that can be mutated to cause FHC. During the past year we have demonstrated that this is the cardiac myosin heavy-chain gene(s).

Chromosomal localization of the disease locus suggested candidate genes that might be responsible for FHC. In particular, cardiac myosin heavy-chain genes (α and β) are located on chromosome 14 band q11–13. During the past year we defined the structure of these genes in affected individuals from two unrelated families (A and B). Affected members of family A have a point mutation in exon 13 of their β gene that changes an arginine residue to a glutamine residue. Because this arginine residue is highly conserved during evolution and because no normal individuals have this mutation, we suggested that this point mutation is responsible for the disease in affected members of this family.

Affected members of family B have an unusual cardiac myosin heavy-chain gene in their genomes. These individuals have a third cardiac myosin heavy-chain gene on one chromosome. This third gene is a hybrid derived from exons 1–27 of the α gene, and the remainder of the gene is derived from the β gene. This hybrid gene is not found in the genomes of unaffected individuals. From these studies we conclude that mutation of these genes can cause FHC.

We have also demonstrated that FHC is a genetically heterogeneous disorder, by studying the inheritance of FHC in two other unrelated families (C and D). In these families the disease is not genetically linked to the cardiac myosin heavy-chain genes on chromosome 14. We propose that their disease is due to defects in other genes that encode products that either affect the expression of cardiac myosin heavy chains or that interact with cardiac myosin heavy-chain genes to produce functional cardiac muscle. We hope that these studies will lead to a better understanding of how cardiac muscle functions and how defects in cardiac muscle proteins lead to the clinical symptoms found in FHC patients.

Computational Neurobiology of the Hippocampus

Terrence J. Sejnowski, Ph.D.—Investigator

Dr. Sejnowski is also Professor at the Salk Institute 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 the 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 was chosen to give this year's Messenger Lectures at Cornell University.

ALTHOUGH there has been an explosion of discoveries over the last several decades concerning the structure of the brain at the cellular and molecular levels, we do not yet understand how the nervous system enables us to see and hear, to learn skills and remember events, to plan actions and make decisions. The long-range goal of this laboratory is to explain how neural systems effect these complex behaviors. Our general approach is to use what we know about the structure and function of identified neurons and neuronal networks to construct computational models at several levels of investigation.

At the biophysical level, the computational mechanisms are based on chemical and electrical signals within and between neurons. Among the more important of these are the signals at synapses, which carry information between neurons. On cortical pyramidal neurons, most of the synapses occur on thorn-like structures protruding from the dendritic shaft, called dendritic spines. The function of these spines is unknown, but is likely to be important because they are found on neurons in most species.

We have recently completed a study of the effectiveness of inhibitory synapses on dendritic spines. About 10 percent of the spines on cortical pyramidal neurons have both excitatory and inhibitory inputs. It had been thought that this arrangement allowed the inhibitory synapse to "veto" the excitatory one when they were simultaneously activated. In our computer simulations of these events, however, we have found that the inhibitory synapses were only effective at reducing the magnitude of the excitatory postsynaptic potential (EPSP) under very restricted circumstances.

The reason is that the volume of a spine is so small that even tiny conductance changes in the membrane allow enough ions to enter and leave the spine to change the intracellular concentrations significantly. This would tend to reduce the forces driving these ions across the membrane and thus reduce the dependent synaptic currents. For our simulations, we infer that, to be effective, the inhibitory synapses found on cortical spines must be mediated by K^+ through $GABA_B$ receptors.

At the neuronal level, we are exploring with computational models the effects of mechanisms that couple the synaptic signals in the dendrites to the spike-triggering region of the neuron, which is located near the soma. According to the traditional view of dendritic processing, current injected into dendrites from synaptic activity is passively conveyed to the cell body by the dendrites' cable properties. In recent years voltage-dependent calcium currents have been identified in dendrites that could boost the coupling between synaptic activity and spike generation. This is particularly interesting because an increase in the excitability of the neuron often accompanies the long-term potentiation (LTP) of excitatory synapses, a rapid and persistent elevation of the EPSP.

Could the increase in excitability seen during LTP result from a change in voltage-dependent calcium currents? To test this hypothesis, we used simulations of intradendritic LTP experiments. The shape of our simulated neuron was based on the detailed branching pattern of a hippocampal neuron that had been injected with a dye and reconstructed in a computer. We found that a small addition of voltage-sensitive calcium channels to the dendritic membrane made a previously subthreshold input suprathreshold, with no significant change in the EPSP.

Our most recent work concerns the effect of these changes in dendritic excitability on LTP specificity. We simulated two sets of synaptic inputs, one stimulated and the other control. Potentiation at the stimulated input had little effect on the control input if the latter was electrically closer to the cell body, or if the two inputs were segregated on different primary dendritic branches. Specificity is degraded if the control input is farther away than the stimulated. These results show that modulation of dendritic excitability is a plausible mechanism for EPSP-spike potentiation.

We are testing this computational model of

EPSP-spike potentiation in two ways. First, we have looked for experimental conditions that would favor EPSP-spike potentiation over synaptic potentiation. Second, we are undertaking direct measurements of intracellular free calcium in hippocampal neurons. The properties of the hippocampal network and its capacity for information storage depend on the properties of the synapses and dendrites. It is therefore important to determine these properties before models of the hippocampus are attempted.

LTP can be induced at some synapses by simultaneous postsynaptic depolarization and binding of glutamate to the NMDA (*N*-methyl-D-aspartate) subtype of glutamate receptor. This is usually accomplished by stimulating the synapses at a high frequency. Is it possible to induce LTP if the dendrite is instead depolarized by invasion from the cell body during an action potential? We have tested this hypothesis in the hippocampus by pairing antidromically elicited action potentials with low-frequency synaptic stimulation. By itself, neither the antidromic spikes nor the synaptic stimulation was sufficient to cause LTP. When these two stimuli were paired, however, there was potentiation of the population spike, although the EPSP did not seem to be increased consistently. This may be due to a specific potentiation of the EPSP-spike potentiation.

The increase in the population spike upon repeated pairing of synaptic and antidromic stimulation is reversibly blocked by the application of AP5, which is an antagonist of the NMDA receptor. This suggests that the depolarization caused by the action potential does invade the dendrites and is acting through the same mechanisms that occur when the cell body is depolarized by current through a recording electrode. It is interesting that Donald Hebb pointed out in 1949 that the firing of postsynaptic action potentials simultaneously with presynaptic activity might be an essential condition for synaptic plasticity. This is the first suggestion that action potentials might in fact have a role to play in inducing LTP. Much more work needs to be done to confirm this interpretation.

If calcium currents in the dendrites of hippocampal pyramidal neurons are indeed responsible for altering the coupling between synapses and the cell body, then it should be possible to verify this directly by measuring the intracellular calcium ion concentration. Calcium ions are known to be central to mechanisms of plasticity in the hippocampus and other regions of the brain, but little is known of the time course or spatial distribution of these changes. Calcium-sensitive fluorescent dyes are first injected into single pyramidal neurons in hippocampal slices. The changes in Ca^{2+} concentration in both the dendrites and soma of the neuron are then monitored with a confocal fluorescence microscope, which allows these changes to be detected with high spatial and temporal resolution. The Ca^{2+} elevations seen in the dendrites were faster than those seen in the soma, but of shorter duration. This would be expected, as their volume is much smaller. We are continuing to develop these techniques and increase the sensitivity of our recordings so that we may be able to correlate physiological changes with changes in the spatial distribution of intracellular calcium.

The projections between areas of the hippocampus have a complex three-dimensional geometry that is difficult to delineate. Anatomists have studied them by injecting tracer into one area and serially reconstructing the terminal arborization of filled neurons. In collaboration with David Amaral of the Salk Institute, we have examined such a data set from injections in area CA3 projecting into CA1 of the rat hippocampus. We then used a mathematical model of these projections to reconstruct a three-dimensional map of the pathway.

By this technique we have uncovered an aspect of hippocampal structure previously unseen: a series of stripe-like modulations along the longitudinal axis. We are now attempting to demonstrate these modulations in direct experiments. We intend to create a realistic model of neuronal networks in the hippocampus from our reconstructed pathways.

The Mammalian Sex Chromosomes

Larry J. Shapiro, M.D.—Investigator

Dr. Shapiro is also Professor of Pediatrics and Biological Chemistry at the University of California, Los Angeles. He received his M.D. degree from Washington University, St. Louis, where he also completed clinical training in pediatrics. He received postdoctoral research training at the NIH. Dr. Shapiro has received many honors for his research and is currently President of the Society for Pediatric Research.

SEX determination is a developmental process of substantial fundamental and practical interest. In a number of vertebrates, sex is determined by environmental factors. For example, some reptiles and amphibians have a temperature-dependent sex-determining mechanism. In various fishes, sex is determined through social cues and interactions. In a number of nonvertebrate species, the somatic sexual phenotype and the actual gonadal sex can be dissociated. In mammals, however, the presence or absence of a Y chromosome is the absolute determinant of sexual differentiation. The presence of one or more Y-encoded genes is necessary and sufficient to initiate differentiation of the embryonic gonad into a testis. Following this commitment, a variety of secondary events result in the elaboration of hormones that produce the full male phenotype.

It is generally believed that mammalian X and Y chromosomes have evolved from a common ancestral homologous chromosome pair. However, a gene or genes must exist on the Y chromosome that does not have a functional counterpart on the X to effect testicular differentiation. To ensure this state of affairs, there must have been a suppression of genetic recombination between the ancestral X and Y chromosomes during evolution. Such suppression likely occurred through the introduction of gross rearrangements in either the X or Y chromosome to make possible the subsequent divergence of DNA sequences.

In modern-day sex chromosomes, however, several areas of persistent X-Y homology can be identified. The first of these is a region approximately 3 megabases long in which absolute sequence identity between the X and Y chromosome distal short arms has been demonstrated. This region facilitates pairing and appropriate segregation of the X and Y chromosomes during male meiosis. The frequent meiotic recombination occurring between the X and Y within this segment ensures the homogenization of DNA sequences in this region.

We have documented the role of this segment of the X and the Y chromosomes in sex chromosome pairing through studies of a patient who had a small deletion of the needed pairing sequences from his X chromosome. As a result of this abnormality, his sex chromosomes were unable to align with each other during meiosis. It is possible that more subtle abnormalities of this pairing region result in some cases of male sterility or predispose to improper sex chromosome segregation and consequent X or Y aneuploidy.

We and others have identified other regions of X-Y chromosome similarity wherein sequences are not absolutely identical. Through the mapping and sequencing of these regions and comparative studies in a number of primate species, we have developed a model for the recent evolution and divergence of the X and Y chromosomes. This model suggests that at least one pericentric inversion of the Y chromosome has occurred, probably within the past 40 to 60 million years of primate evolution. As a consequence, some DNA sequences are represented on the long arm of the Y chromosome that have homologues on the short arm of the X.

Many of the sequences on the Y chromosome appear to have undergone degeneration and to have accumulated mutations that render them molecular fossils with no apparent function. However, we have recently studied the structure and expression of the human amelogenin genes, which are located in the same general area. The amelogenin X and Y genes are approximately 90 percent similar in their DNA sequences, but are predicted to encode proteins of varying amino acid composition. Through studies of RNA production in developing tooth buds, we have shown that both the X and Y genes are functional and can be expressed in males, leading to the production of two different amelogenin proteins. Thus not all of the sequences involved in the Y chromosome pericentric inversion have undergone degeneration.

Another practical aspect of the X-Y homology is that it can occasionally be the site of aberrant recombinational events producing X/Y translocations. We have studied a number of these patients in some detail and have shown that this cytogenetic abnormality is in fact the result of

unequal exchange within this region of homology between the X and Y chromosomes.

Another area of substantial interest in the laboratory relates to deletions in X chromosome–encoded genes. Duchenne muscular dystrophy is a human genetic disorder resulting from abnormalities in an X-encoded gene called *dystrophin*. The majority of patients with Duchenne dystrophy have deletions of part or all of the *dystrophin* gene.

Patients with X-linked ichthyosis, a heritable skin disorder, have deficient activity of the enzyme steroid sulfatase (STS). This gene is encoded in 150 kilobases of X chromosomal DNA, and in 90 percent of patients with this relatively frequent cutaneous disorder, the entire gene is deleted. Our studies have established that such deletions result from recurrent *de novo* events on the X chromosome. These involve the unequal exchange of DNA catalyzed by a low-copy-number repetitive element that flanks the STS gene.

Thus recombination between similar but nonidentical elements appears to be responsible for the genesis of these common deletional events. We are currently trying to determine whether this process occurs with a high degree of frequency elsewhere within the genome.

The phenomenon of X chromosome inactivation is another area of interest in our laboratory. Since all normal female mammals have two X chromosomes and males have only a single X, there is a fundamental dosage inequity. This is normalized for most genes by functional inactivation of one of the X's in female somatic cells very early in embryogenesis. Previously, we have shown that X chromosome inactivation is integrally related to DNA methylation. Currently we are investigating a model of X chromosome inactivation *in vitro* involving the use of differentiating teratocarcinoma cells. We hope to identify proteins and other gene products that are important in counting the number of X chromosomes present and in initiating the X inactivation process.

There is also a small group of genes that appear to escape the process of X inactivation even when they are situated on otherwise inactive X chromosomes. Our previous studies have shown that these genes escape inactivation in a position-independent fashion. We are currently studying the regulatory signals involved in X inactivation and those unique features that render some genes resistant to this phenomenon. We have cloned and characterized the promoter of the STS gene as an example of this class of genes and have found what appears to be a novel promoter structure.

We are trying to identify other genes that escape X inactivation so that we can compare regulatory regions between them and identify DNA sequences that are critical for initiating and propagating X inactivation signals along the chromosome. The studies described should provide a better understanding of the evolution of the sex chromosomes and insight into the pathogenesis of the many human disorders that are due to mutations or structural and numerical abnormalities of these chromosomes.

Adenovirus as a Model for Oncogenesis and 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 trained as a postdoctoral fellow with Paul Berg at Stanford University. Before joining the faculty at 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 in nature. Humans are first infected when very young. Generally the infection results in cold-like symptoms, and the episode resolves without complication. Injected into a rat or hamster, some human adenoviruses induce a variety of benign or malignant tumors. Adenoviruses are grouped in the class of DNA tumor viruses, since they are tumorigenic under certain conditions and contain DNA chromosomes.

Human adenoviruses can be propagated easily in cultured human 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 virus-coded protein shells to generate virus particles. Since viral gene expression is tightly regulated and occurs at high levels compared with that of most cellular genes, adenoviruses are a useful probe for the study of gene expression control.

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 at the level of transcription. Expression of these viral genes can be induced by treatment of cells with cAMP.

Most viral genes activated by the E1A protein contain a small sequence termed a cAMP response element (CRE) within the DNA sequences of the transcriptional control region. Cyclic AMP is a key player in a regulatory cascade that can induce transcription of a variety of cellular genes. Increased cAMP levels lead to activation of the CRE-binding (CREB) protein, which in turn binds to the CRE sequence.

The E1A gene product and cAMP act in synergy to induce transcription of target genes. We have not detected a change in the level of CREB activity under induced conditions. However, a second transcription factor, AP-1, is strongly induced. AP-1 binds to a DNA sequence termed a TPA response element (TRE). The TRE sequence differs from the CRE by only 1 base pair, and AP-1 binds almost as well to CRE as to TRE sequences. Thus our current hypothesis is that the E1A gene product plus cAMP acts to raise the activity level of AP-1, which binds to both CRE and TRE sequences and thus contributes to the activation of viral genes.

The induction of AP-1 activity occurs at two levels. The first is a rapid modification of the factor, causing it to display altered physical properties. The second level requires active transcription of the products of two proto-oncogenes, c-*fos* and *junB*. Together these products constitute one form of AP-1. It is likely that the modified form of AP-1 not only induces expression of adenovirus genes but also of its own constituents, activating an autoregulatory loop. We are now working to identify the constituents of the modified AP-1 activity and to delineate the role that the E1A product plays in the modification process.

The induction of AP-1 activity by E1A plus cAMP is transient. A second protein, encoded by the adenovirus E4 gene, causes AP-1 activity to return to basal levels several hours after treatment of infected cells with cAMP. Thus the induction of AP-1 is tightly regulated in the infected cell. The E4 protein also causes a reduction in the level of phosphorylation of the c-*fos* component of AP-1. We are presently working to determine whether this change in phosphorylation is functionally related to the reduction in AP-1 activity and to elucidate the mechanism by which the reduction occurs.

The E1A protein, in addition to inducing a positively acting transcription factor, AP-1, can also inhibit the activity of a negatively acting factor, which 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 adenovirus E1A protein activates expression of the P5 con-

trol region. In the absence of E1A protein, YY-1 represses P5 transcription. In its presence the repression is relieved, and the control region becomes transcriptionally active.

To study the YY-1 factor, we prepared some from cultured human cells and obtained 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. Protein was expressed from the clone and shown to bind specifically to the YY-1 recognition site. Sequence analysis of the clone revealed that YY-1 is a 68-kDa protein with a zinc finger DNA-binding motif.

We are presently studying the mechanisms by which YY-1 represses transcription and by which the adenovirus E1A protein relieves the repression. We also wish to determine whether E1A protein mediates its effects on the positive-acting AP-1 and the negative-acting YY-1 factors through the same or different mechanisms. That is, does the E1A protein mediate just one biochemical reaction that affects AP-1 and YY-1 differently, or does it carry out multiple, physiologically distinct biochemical processes?

In addition to the E1A protein, which functions to induce transcription, we have studied the adenovirus 55-kDa protein induced by E1B. This protein also controls the accumulation of mRNAs, but it acts after transcription. It simultaneously blocks the accumulation of cellular mRNAs and enhances accumulation of viral mRNAs. Further studies indicated that the protein's discrimination of viral from cellular mRNAs is based on the site of synthesis. If an RNA is transcribed from the viral chromosome, its accumulation is enhanced; if it is transcribed from a cellular chromosome, its accumulation is blocked.

We determined the localization of the E1B 55-kDa protein within the infected cell, using a technique termed immunogold electron microscopy. The protein was localized to the periphery of spherical bodies previously shown to be sites of viral DNA replication and transcription.

Localization of the E1B protein at the replication-transcription centers is consistent with its ability to enhance accumulation of viral mRNAs. Its location makes it available to these mRNAs shortly after their synthesis and processing. But how does this localization prevent cytoplasmic accumulation of host cell mRNAs? It seems likely that the protein interacts with and draws to the periphery of viral replication-transcription centers a cellular factor that functions to move mRNAs from their site of synthesis and processing to the nuclear pore for transport out of the nucleus. Relocation of a cellular factor required for intranuclear movement of RNAs could explain how the E1B protein can inhibit accumulation of RNAs transcribed from cellular locations and simultaneously stimulate accumulation of transcripts derived from the viral chromosome. We are presently searching for cellular factors that interact with this protein.

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. White cells, unlike the continuously circulating red cells, migrate from small blood vessels into other organs, where they participate in the immune and inflammatory reactions that characterize host defense. Different classes of white cells carry out specialized functions: macrophages and granulocytes ingest and kill microorganisms, whereas lymphocytes recognize foreign antigens and produce antibodies to combat them.

To maintain constant numbers of cells in the circulation, the process of blood cell development (hematopoiesis) must be subject to exquisitely sensitive regulatory controls. Diseases affecting blood cell production include those in which insufficient numbers of red and white cells enter the circulation (anemias and leukopenias); those in which the numbers of normal white cells are increased (leukocytosis, seen in systemic infections); and those in which cells are abnormal (dysplasias) or cancerous (leukemias).

Hematopoiesis is regulated by a group of protein growth factors, collectively termed colony-stimulating factors (CSFs) or interleukins. These polypeptides are produced by resident stromal cells in the bone marrow as well as by circulating blood cells (the term interleukin literally meaning *between white blood cells*). First recognized through their ability to stimulate immature bone marrow–derived (myeloid) precursor cells to form colonies composed of differentiated blood cell elements, CSFs were named for the types of colonies they produced. For example, M-CSF (or CSF-1) specifically induces macrophage colonies, G-CSF supports granulocyte development, and GM-CSF stimulates the proliferation and maturation of both cell types.

CSFs, now produced in previously unobtainable quantities by genetic engineering techniques, have become part of the clinical armamentarium and have proved particularly useful in reversing certain bone marrow failures and in heightening impaired host defenses against infection.

Signal Transduction by the CSF-1 Receptor

CSF-1 (M-CSF) supports the growth, maturation, and survival of macrophage precursors in the bone marrow and potentiates the effector functions of mature macrophages during the inflammatory response. Its diverse physiologic actions are mediated through its binding to the CSF-1 receptor (CSF-1R), a cell surface glycoprotein. The receptor consists of an extracellular ligand-binding portion, joined through a single membrane-spanning segment to an intracellular kinase domain capable of phosphorylating other cellular proteins on tyrosine residues. Receptor-mediated phosphorylation modifies the biochemical behavior of several effector proteins, which then relay signals to the cell nucleus that effect changes in gene expression, DNA synthesis, and cell division.

CSF-1 induces dimerization of its receptor at the cell surface, activating CSF-1R protein kinase activity and leading to the cross-phosphorylation of receptor subunits on tyrosine. The autophosphorylation of CSF-1R triggers its association with other cellular enzymes, whose biochemical activities in signal transduction are modified by their binding to the receptor, their phosphorylation on tyrosine, or both. Mutant CSF-1R molecules lacking particular autophosphorylation sites are defective in some aspects of signaling but not others, suggesting that the combinatorial actions of enzymes that associate with the receptor can in part determine the specificity of the CSF-1 response in different cell types.

CSF-1R is normally restricted in its expression to macrophages and their precursors; it is not detected on cells of other hematopoietic lineages. However, introduction of the gene encoding CSF-1R into cells that depend upon other growth factors enables them to respond to CSF-1. Expression of CSF-1R in immature myeloid cells allows CSF-1 to replace their interleukin-3 requirements for growth and viability.

By contrast, transduction of CSF-1R into immature pre–B lymphoid cells not only relieves their normal dependence on interleukin-7, but can reprogram the fate of responding cells, enabling them to trans-differentiate to mature macrophages in response to CSF-1. Physiologic targets of the CSF-1R kinase must therefore be expressed more ubiquitously than the receptor itself and, depending on cell context, can modify genetic programs governing both cell proliferation and fate.

Transforming Potential of CSF-1R

CSF-1R is encoded by the *FMS* proto-oncogene, which can be converted by mutation to an oncogene that induces tumor formation. Mutations in the extracellular domain of CSF-1R activate the receptor kinase in the absence of CSF-1, leading to sustained, unregulated signals for cell growth. By mutagenizing segments of human *FMS* and screening "libraries" of mutated genes for their ability to induce cell transformation, we localized particular subdomains within the CSF-1R extracellular domain where "activating mutations" occur. Postulating that similar genetic alterations might contribute to leukemogenesis, we are using polymerase chain reaction techniques to assay for the presence of activating mutations in the *FMS* genes of myeloid tumor cells.

CSF-1-Responsive Genes Required for Cell Division

CSF-1R regulates genes that collectively govern macrophage growth, survival, differentiation, and effector functions. When proliferating macrophages complete cell division (mitosis), they enter a 10- to 12-hour gap phase (G1) in their cell cycle before synthesis of chromosomal DNA is reinitiated (S phase). CSF-1 is required throughout G1 for the cells to enter S phase, but once they begin to replicate DNA, they can complete S phase and divide in the absence of the growth factor. The requirement for persistent CSF-1R-mediated signals throughout G1 implies that the expression of growth factor–responsive genes is temporally regulated over a relatively long period. Conceivably, CSF-1-responsive gene products synthesized early in G1 feed back to modulate receptor-mediated signals, so that the genes that commit cells to DNA synthesis are ultimately induced.

Genetic data accumulated through studies of yeasts indicate that a class of cell cycle genes called G1 cyclins might act to prime cells for DNA synthesis. We recently isolated genes from CSF-1-stimulated mouse macrophages that have predicted structural motifs reminiscent of known cyclins and whose expression is regulated in G1. Related genes are expressed in other cell types, suggesting the presence of a novel cyclin gene family. If we can demonstrate that expression of these cyclins governs G1 progression in mammalian cells, we may be mechanistically able to link early steps in CSF-1-mediated signal transduction with later events required for DNA synthesis.

Modulation of Internal Calcium and Synaptic Function by Neurotransmitters

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 spent two years as a postdoctoral fellow in the laboratory of David Colquhoun at University College, London, and one year with Philippe Ascher at the Ecole Normale Supérieure in Paris, studying 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 ability of neurons to regulate and alter the strength of their synaptic connections is thought to play a crucial role during learning and development. Changes in synaptic function can be brief, lasting seconds to minutes, or prolonged, lasting days to months. One important mechanism by which neurons regulate synaptic transmission is by regulating the amount of neurotransmitter released from the presynaptic cell. Since this release is triggered by the influx of calcium ions into the presynaptic terminal, it has been widely proposed that modulation of release results from the modulation of presynaptic calcium levels. However, there is little direct evidence to support this hypothesis.

For the past several years we have focused our attention on the up- and down-modulation of neurotransmitter release from the terminals of *Aplysia* sensory neurons in response to the modulatory transmitter serotonin (5-HT) and the neuropeptide FMRFamide, respectively. Electrophysiological recording techniques have shown that 5-HT increases action potential duration by closing a particular potassium channel (the S channel) and increases an inward calcium current that is sensitive to block by dihydropyridine drugs (the L-type calcium channel). In contrast, FMRFamide increases the opening of the S potassium channel, decreasing action potential duration, and decreases current flow through a class of calcium channels not blocked by dihydropyridines (the N-type channels).

In a recent study, Brian Edmonds, Nicolas Dale, Marc Klein, and Eric Kandel (HHMI, Columbia University) have shown that only calcium that enters through the N-type calcium channels is able to trigger transmitter release. This conclusion was based on the finding that application of dihydropyridines to block the L-type calcium channels has little effect on transmitter release. Thus presynaptic facilitation with 5-HT is thought to result from an indirect increase in calcium influx via the N channels into the sensory neuron presynaptic terminals due to the increase in action potential duration (which allows the N-type calcium channels to remain open longer). Presynaptic inhibition with FMRFamide is thought to result from a direct decrease of calcium influx via the N-type calcium channels.

Do 5-HT and FMRFamide modulate calcium influx during an action potential in the presynaptic terminals of sensory neurons? To what extent is such modulation due to changes in calcium influx via N-type as opposed to L-type calcium channels? Why doesn't calcium entry via the L-type channels contribute to transmitter release?

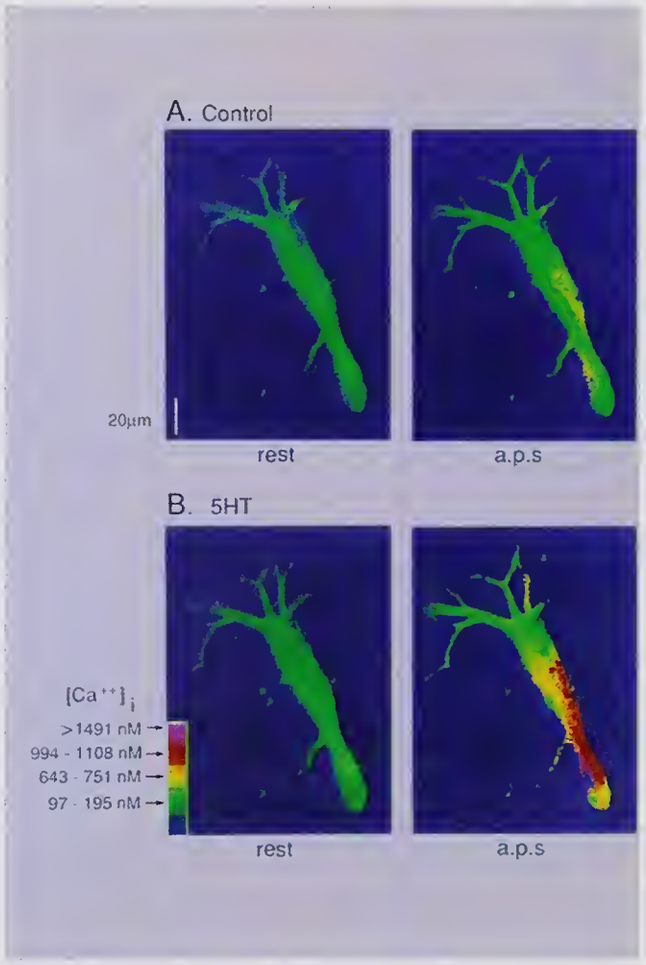
To address such questions we have used the fluorescent dye fura-2, developed by Roger Tsien (HHMI, University of California, San Diego), which binds calcium and reports intracellular calcium concentration, to study the effects of 5-HT and FMRFamide on intracellular calcium levels in response to action potential stimuli. In a previous study we showed that the action potential–induced calcium transient was increased by 5-HT and decreased by FMRFamide in sensory cell bodies, growth cones, and neurites. Over the past year we have investigated whether such modulatory changes occur at presynaptic terminals, by coculturing sensory neurons with postsynaptic motor neurons. Pseudocolor images of intracellular calcium levels were obtained at presumed presynaptic sensory neuron terminals at regions of contact between sensory and motor neurons. The intracellular calcium transient induced by action potentials was markedly enhanced by 5-HT. These results show that presynaptic calcium transients are modulated at defined regions of contact between sensory and motor neurons.

Why is it that calcium influx via the L-type channels does not contribute to transmitter release? One possibility is that during an action potential, the L-type channels contribute relatively little to total calcium influx. Alternatively, these channels may carry substantial amounts of calcium into the cell, but the channels may be localized in regions of the cell far from the site of synaptic contacts.

Measurements of calcium transients in response to action potentials show that, in the absence of 5-HT, dihydropyridines have a relatively small effect, reducing calcium entry by less than 10 percent. Thus the lack of effect of dihydropyridines under normal conditions could simply be explained by their small contribution to total calcium influx. However, in the presence of 5-HT, application of dihydropyridines causes a substantial decrease in the amount of calcium influx, without affecting transmitter release. This leads us to conclude that the L-type calcium channels must be segregated from release sites.

How much of the increase in the calcium transient with 5-HT is due to modulation of the L-type calcium channels and how much is due to the increased calcium influx via the N-type calcium channels as a consequence of the increase in action potential duration? To answer this question we compared the increase in the calcium transient in response to 5-HT in the absence and presence of dihydropyridines. About two-thirds of the increase is due to the increase in the L current, while one-third is due to calcium influx via the N-type channels, which presumably contribute to the increase in transmitter release.

These results provide direct evidence in support of the hypothesis that alterations in calcium influx at presynaptic terminals is involved in modulating the strength of synaptic transmission. However, the results also raise several questions currently under investigation. To what extent does the change in calcium account quantitatively for the change in transmitter release? Can we visualize local clusters of L-type and N-type calcium channels and show that only the N-type channels are present at presynaptic release sites? To what extent are long-term changes in synaptic transmission due to changes in calcium entry? What role do the L-type channels play in cellular physiology?



Pseudocolor image of intracellular calcium concentration in a presynaptic process of an Aplysia *sensory neuron. A: Left panel shows that, at rest, resting calcium is low. After stimulating a brief train of action potentials, there is a rise in intracellular calcium.*

B: After application of the facilitatory transmitter serotonin (5-HT), the same train of action potentials gives rise to a much larger increase in calcium. Scale applies to all panels.

Research of Steven A. Siegelbaum.

Chemical Details 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 the 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.

OUR goal is to understand how the cell responds to regulatory signals. We have focused on two processes: the transmission of signals across the membrane, and the regulation of gene expression. The spirit of the work is reductionist, rather than descriptive; we aim to understand the key steps of these processes in terms of basic physical and chemical principles. To do this, we must visualize the detailed structure of the relevant macromolecules and the complexes they form with other molecules. The best way to do this is to crystallize these complexes and determine their structure by x-ray diffraction.

The stereochemical underpinning for the transmission of regulatory signals through cell membranes is a broad and varied subject. We have begun by focusing on the clinically important question of inflammation. When the signals are transmitted, phospholipases are stimulated to attack phospholipids, the main substance of membranes. The breakdown products are used to supply precursors or signals for subsequent reactions inside the cell. The enzyme phospholipase A_2 (PLA_2) is thought to be responsible for producing arachidonic acid, which is the precursor of most of the small compounds that mediate inflammation.

Recently we defined the elusive mechanism by which PLA_2 hydrolyzes phospholipids and releases arachidonic acid. We worked out the mechanism by solving the crystal structures of enzyme-inhibitor complexes in which the inhibitor was designed—by Michael Gelb of the University of Washington—to simulate the crucial catalytic intermediate, or "transition," state. Enzymes can speed reactions by stabilizing such states, and these crystal structures show clearly how PLA_2 does this. We now have three crystallographically independent structures of these complexes, and all show the same characteristic relationship between the enzyme's active surface and the transition-state analogue.

These structures have additional and special interest to a wide scientific audience. We have defined, in atomic detail, the mechanism of a calcium ion's regulatory activity at the cell surface. Moreover, we have shown how a soluble protein interacts with the face of the cell membrane. Besides answering general scientific questions, these studies afford an immediate practical gain. Our structures may provide a basis for the rational design of therapeutic agents that block an inappropriate or exceptionally severe inflammatory response. To this end, we have recently obtained suitable crystals of PLA_2 found in the inflamed joints of patients with acute rheumatoid arthritis and the serum of women with toxic shock syndrome. The crystal structure of this enzyme should provide a starting point for the rational design of drugs aimed at these and related inflammatory diseases.

Ultimately most regulatory signals control the expression of genes, turning some off and others on. Much of this occurs by regulating transcription (the synthesis of messenger RNA). An essential element of transcriptional regulation is to target the regulatory proteins to the genes they are designed to control. For example, the estrogen receptor, which is an activator of transcription, is obviously targeted to different genes than its counterpart, the testosterone receptor. This is usually accomplished by a recognition process whereby the regulatory protein binds to a specific DNA sequence associated with the genes to be controlled. Recognition involves the formation of a strong interaction between the protein and the correct DNA sequence and weak interactions with other DNA. We have focused our attention on the basic chemistry responsible for stabilizing specific protein-DNA interactions.

Our results in this effort are most advanced in the case of the *trp* repressor-operator complex, which regulates the expression of genes responsible for tryptophan biosynthesis in *Escherichia coli*. The structure was refined to unusually high resolution (1.9 Å). This refinement, coupled to the fact that the same structure is found in four independent representations, provides us with a model of exceptional clarity and detail.

The chemistry of the interface was unex-

pected. Most of the direct interactions are hydrogen bonds that appear to lash the protein to the sugar phosphate backbone on both sides of the DNA's major groove. In some instances a single phosphate accepts four hydrogen bonds. An intimate specific fit is partly accomplished by a deformation of the DNA, which we believe to be allowed specifically by the target's sequence at little or no cost in internal energy. We refer to this sequence-dependent deformability as "indirect read-out" of the sequence.

Most surprising is the array of fixed water molecules (20 per complex) that help mediate the specific interaction. Ten of these are also found in the uncomplexed repressor protein and therefore represent noncovalent extensions of the protein's "recognizable" surface. Four of the ten fixed water molecules provide six highly specific hydrogen bonds to bases that define the *trp* operator as the target for the *trp* repressor. Recent mutational studies designed to test the functional significance of these specific water-mediated contacts support the idea that water molecules can mediate molecular recognition. The obligatory involvement of water in molecular recognition has also been reported recently for highly specific interactions of proteins with sugars and RNA molecules.

Our most recent work has focused on understanding DNA recognition in higher cells. We are trying to determine the crystal structures of a variety of regulatory protein-DNA complexes that control metabolism, growth, differentiation, and the expression of viral-specific genes. Some of these complexes closely resemble the products of oncogenes.

We have just determined the crystal structure of the DNA-binding domain of the glucocorticoid (cortisol) receptor in a complex with its DNA target. The ability of this protein and each of its close homologues to bind to its specific target sequence lies at the heart of the selective expression of genes in response to each of the different steroid hormones. Despite their discrimination in targeting the correct regulatory sequence, the DNA-binding domains of all the steroid receptor "superfamily" (thyroid hormone included) have a highly conserved sequence and presumably very similar structures. We found that these DNA-binding domains contain two intriguing substructures called zinc fingers, found in many genetic regulatory proteins. Mutagenesis experiments done elsewhere have indicated where to focus our attention, and we can now account for the mechanism by which the glucocorticoid receptor can discriminate between its target DNA and those of the estrogen and thyroid receptors.

Work on other systems is not quite so far along. We hope to grow better DNA-protein crystals containing the binding domain of the "leucine zipper proteins," another important type of transcription factor first characterized by Steven McKnight (HHMI at the Carnegie Institution). Work also continues with Laimonis Laimins (HHMI at the University of Chicago) on strongly diffracting crystals of the specific DNA complex of a regulator protein, E2, which can trigger the transforming activity of the potentially oncogenic papilloma virus.

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 does the regulatory protein on binding near its target gene 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.

Current research focuses on the regulatory functions of Oct-2 in B cell development, as well as the mechanism underlying lineage-restricted expression. Oct-2 is likely to represent a member of a regulatory gene hierarchy that controls the development of B cells from pluripotent hematopoietic stem cells. By studying the regulation of Oct-2 expression, we can identify and isolate genes interacting with Oct-2 in the hierarchy. The murine gene encoding Oct-2 has been isolated in the laboratory and is being analyzed for regulatory sequences that control its expression in developing B cells.

The activity of the Oct-2 gene is regulated during B cell differentiation. The Oct-2 gene is expressed at low levels in pre–B cells. Oct-2 expression is transcriptionally up-regulated 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, another regulatory protein that has been implicated in the control of the activity of the κ locus. We propose 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. Our results suggest that NF-κB may also be involved in regulating Oct-2 transcription during pre–B cell differentiation. Therefore this transition in B cell ontogeny may be controlled by a cascade of transcriptional regulators.

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 an organism from infection by a variety of means. Thus an effective immune response depends on a properly functioning hematopoietic pathway.

When defects arise in this 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 need to explain at the molecular level the regulation of the hematopoietic pathway.

The primary objective of our research is to identify and characterize regulatory proteins for one stage of hematopoiesis: the maturation pathway for B lymphocytes. These cells produce antibody molecules and therefore play a central role in an immune response. In numerous laboratories a variety of approaches have been employed to define molecules that regulate B cell maturation.

One approach has been to identify components of the bone marrow microenvironment (which surrounds precursor B cells) essential for cell growth and differentiation. One such component is a soluble growth factor called interleukin-7, which is secreted by bone marrow–derived cells and promotes proliferation of specific pre–B cell types. Additional growth factors and stromal cell membrane proteins have been identified that may promote further stages of B cell development.

A second approach has been to characterize genes that obstruct the differentiation process. These genes include several viral oncogenes and their cellular homologues. The v-*abl* oncogene, for example, is leukemogenic in specific pre–B cell types, preventing them from further differentiation. The c-*myc* oncogene also facilitates pre–B cell transformation, and one study has shown that c-*myc* transformation was prevented by artificially altering the B cell differentiation pathway. Further studies of the structure and function of these oncogenes are providing insight into the control of differentiation.

We are utilizing a third approach to define B cell regulatory molecules. We are identifying proteins that directly activate or inactivate genes expressed at specific stages of differentiation. The regulation of gene expression appears usually to be carried out by modulation of the rate of initiating transcription (RNA synthesis). Transcriptional analysis as a means of studying B cell differentiation has already been employed successfully and extensively with the genes that encode the antibody, or immunoglobulin (Ig), molecules. In fact, transcription of these Ig genes has been studied in more detail than that of almost any other cellular gene.

These studies have revealed at least two important regulators of B cell differentiation. Both regulatory proteins, OCT-2 and NF-κB, bind to DNA sequences in the control regions for Ig genes and play a role in specific activation of Ig RNA synthesis in developing B cells. Besides these two proteins, several other DNA-binding proteins, currently less well characterized, have been identified that also may play a role in regulating Ig transcription.

The analysis of Ig transcription and the identification of OCT-2 and NF-κB have contributed significantly to our understanding of B cell differentiation. Surprisingly, however, there has been little effort toward characterizing the transcriptional regulation of several other genes known to be expressed specifically in early B cells. We therefore have chosen to extend the analysis of B cell differentiation by focusing on the transcriptional regulation of a new gene that is regulated in a precise fashion during B cell ontogeny: the terminal deoxynucleotidyltransferase (TdT) gene. This gene encodes a DNA polymerase that appears to play a role in generating diversity within the antibody molecules.

We chose to analyze TdT expression for four reasons. First, the gene is tightly regulated, being expressed exclusively in immature lymphocytes.

Many other genes expressed in early lymphocytes are also expressed in other tissues. Second, unlike the Ig genes, which are permanently activated during B cell differentiation, the TdT gene is only transiently activated and then remains off in the mature cell. This expression pattern provides the opportunity to analyze mechanisms for transcriptional activation as well as inactivation during lymphopoiesis. Third, the TdT gene is not only expressed in early B cells, but also in developing T lymphocytes, which are responsible for the cell-mediated immune response. It will therefore be interesting to analyze common regulatory proteins and pathways for these two distinct lymphoid lineages. And finally, until the recent isolation of two new genes, called RAG-1 and RAG-2, TdT was the only gene available whose product was thought to play a specific role in the gene rearrangement process responsible for antibody diversity. Because the rearrangements are likely to be key events that drive B and T lymphocyte differentiation, it is important to investigate the regulation of the rearrangement machinery in order to understand the control of lymphopoiesis.

Our analysis of TdT regulation has suggested, quite surprisingly, that the TdT control mechanisms are fundamentally very different from Ig control mechanisms. Moreover, the TdT transcriptional control mechanisms may be representative of those for a variety of genes expressed in early B and/or T cells.

Both the general structure of the transcriptional control region within the TdT locus (called the promoter) and the lymphocyte-specific DNA-binding proteins for TdT transcription appear to be unique when compared with our knowledge of Ig transcription. The TdT promoter does not contain a TATA box, a common DNA sequence element found in the Ig promoters and in most other promoters that have been characterized in detail. Moreover, the lymphocyte-specificity of the TdT promoter appears to be regulated in part by a novel lymphocyte-specific DNA-binding protein, called LyF-1, rather than by the OCT-2 or NF-κB proteins.

Most importantly, these features of the TdT promoter are found in the promoters for a number of other genes activated in early B and/or T cells. These include the B cell–specific $\lambda 5$ and VpreB genes and the two distinct promoters for the T cell–specific *lck* gene. All of these promoters lack TATA boxes and contain at least two binding sites for LyF-1.

Our studies currently are focused on understanding the two unique qualities of TdT transcription described above: 1) the ability to direct accurate RNA synthesis in the absence of a TATA element and 2) the apparent use of the LyF-1 protein to control the lymphocyte specificity of TdT expression.

In place of the TATA element, which typically is located 30 nucleotides upstream of the transcription start site, the TdT promoter contains a distinct element that overlaps the start site. This element, which we call an initiator (Inr), is like the TATA element in that it is important for promoter function and also pinpoints the RNA start site to a specific nucleotide. We recently have found functional Inr elements in many different genes, even though a precise, consensus DNA sequence has not been detected.

We have also found, through analysis of synthetic promoters, that the TATA element is dominant over the Inr element. In a promoter containing both TATA and Inr, the TATA element directs the location of the transcription start sites. Other studies are providing further insights into the functioning of the Inr element and its relationship to TATA.

The LyF-1 protein appears to play a role in the lymphocyte specificity of TdT transcription. It is found predominantly in lymphoid cells and binds tightly to two sites in the TdT promoter that are important for lymphocyte-specific transcription. We have purified this protein in order to demonstrate that it can bind to promoters for the other lymphocyte-specific genes mentioned above. Moreover, the purified protein is being used to reach one of our primary goals: isolation of the gene encoding LyF-1. This gene will allow us to understand better the mechanism by which LyF-1 regulates TdT expression. Ultimately, our continued analyses of LyF-1 and of TdT expression are likely to broaden our knowledge of the control of hematopoiesis.

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 MIT, before joining the faculty at Baylor. He is a Pew Scholar in the biomedical sciences.

THE major aim of my laboratory is to extend the understanding of early development of the mouse, both by studying cell lineages and by generating mutations that affect the embryo.

In the study of cell lineages, we are particularly interested in determining the origin of germ cells and in establishing the time of allocation of cells to specific lineages once the embryo has been implanted in the uterus. In these studies, embryos are infected at various stages of development with retroviruses transducing the β-galactosidase (βgal) gene of *Escherichia coli*. The gene, used as a tracer, permits histochemical analysis of infected cells and their progeny, which in turn permits the establishment of cell lineage relationships between βgal cells in a clone.

Following implantation, the embryo undergoes a series of cleavages accompanied by extensive cell mixing. Gastrulation begins at day 7 of development with the formation of the primitive streak. We have observed frequent infection of postimplantation embryos with βgal viruses expressing the gene under the control of the viral promoter or of an internal phosphoglycerate kinase 1 promoter. Infection of embryos at day 7 and 8 of development suggests that extensive cell mixing stops at this time, since embryos with single clusters of βgal-expressing cells are readily observed.

The second aspect of our research concerns the identification of developmentally regulated genes in the embryo. Despite the mapping of numerous mutations, both spontaneous or induced by radiation or chemicals, it has been difficult to determine the molecular defect on mouse chromosomes that causes a given phenotype. For this reason, insertion mutagenesis, in which DNA introduced into the germline can cause a mutation by disrupting or affecting the expression of a gene, is an attractive research approach, with the transgene serving as a tag for the molecular cloning of the affected gene. The DNA can be introduced into the embryo by retroviral infection or direct microinjection into the zygote pronuclei, or by using embryonic stem (ES) cells, which can be infected or electroporated and then selected and introduced into embryos to colonize the germline.

Over the past year we have developed a method for generating large numbers of mutations in mice. The method is useful because, on the average, only 1 out of 15 transgenic strains exhibits an overt mutant phenotype. To avoid extensive breeding of useless strains, we devised a screen based on "promoter traps." A reporter gene is placed downstream of a DNA 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 selected ES cells can be used both to trace the activity of the tagged gene, by expression of the reporter, and to mutate the strain.

We have used as reporters both the βgal gene and a fusion protein encoding two enzymes, βgal and neomycin phosphotransferase. This fusion protein, βgeo, allows direct selection for promoter trap events. Thirty-four transgenic lines have now been generated using a retroviral promoter trap vector, and they are being examined both for patterns of expression and for phenotype. Among the first 13 strains analyzed, some have been found in which the promoter trap caused embryo death and some in which no overt phenotype is associated with the mutation. This suggests that a certain proportion of genes are unessential for normal development. To identify the mutated genes, we have started to clone the sites of viral integration.

ES cells can be used also to select for mutations in specific genes. Efforts in the laboratory have focused initially on the gene encoding c-*src*, a proto-oncogene that is a tyrosine kinase. It is thought to play a role in the cell cycle and is highly expressed in neurons and platelets. Constructs designed to delete the c-*src* activity were introduced into ES cells by electroporation, and homologous recombinant clones were isolated by screening with the polymerase chain reaction (PCR) technique. In a given construct, which includes about 8 kilobases of homology, the frequency of such targeted events is about 1 in 100. Southern blot analysis has demonstrated that

these are authentic homologous recombination events.

ES cells have been used to generate germline chimeras. Animals heterozygous for the *src* mutation exhibit no phenotype, whereas homozygous animals survive to birth but are retarded relative to their littermates and die in the first few weeks. Further analysis of these animals reveals an osteoclast defect and resulting osteopetrosis. This result is surprising, for it implies that *src* may not be essential for cell types in which it is most highly expressed.

In many cell types, other tyrosine kinases related to c-*src* may play overlapping roles. We are particularly interested in the c-*yes* and *fyn* genes. To study these, we have generated mutations in ES cells by homologous recombination in c-*yes* and are designing similar experiments with *fyn*. The frequency of homologous recombination at the c-*yes* locus has been about 1 in 30, and the *yes* chimeras will be bred to ascertain contribution to the germline. The resulting mutant mice may be examined for the effect of loss of *yes* and crossed to *src*-deficient mice to examine the interplay between different members of the *src* gene family.

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 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.

EGGS have always fascinated biologists. The union of egg and sperm initiates the complex developmental processes that ultimately produce a new multicellular animal. However, the roles played by egg and sperm 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 sperm nor any other cell can even begin the complex processes that lead to an embryo and ultimately to 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, then that cell would surely be an egg.

Not surprisingly, egg cells are very different in structure as well as in biological capacity from other cells. Most chicken cells weigh less than 1 millionth of an ounce, whereas a single chicken egg makes a nice breakfast. However, in addition to size, eggs diverge in many important ways from other cells. The genes within the egg's chromosomes function differently. These alterations are so extensive that egg chromosomes appear in the microscope quite unlike those in other tissues. As a result of these differences, in eggs the products of many genes accumulate and are stored in special forms that can be utilized at precisely appropriate times during embryonic development. Indeed the presence of so much stored material is one of the reasons for the enormous size of many eggs.

A great deal remains to be learned about the structure of eggs. How many gene products are specially stored there, and what are their specific functions later in development? Is each product located in a particular place? What are the special mechanisms that allow gene products 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 prevented all attempts to describe egg structure in molecular detail, much less to understand the logic that allows this structure to begin development into an even more intricate adult organism. However, during the last 10 years or so, the advent of powerful new methods in molecular biology, such as gene cloning and gene transfer, in combination with *Drosophila* (fruit fly) genetics, has begun to unravel the fascinating secrets stored inside an eggshell.

Using Genetics to Study Eggs

It is now possible to study structures as complex as eggs, because 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, since over 70 years of genetic studies have revealed far more about its genes than those of any other complex organism, including the human. A genetic mutation inactivates one specific gene among the 20,000 to 50,000 in the fruit fly. 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 due to a lack of stored food. By studying what goes wrong with the mutant eggs, biologists can learn much about what the product of that particular gene does normally. For example, in this case it would be concluded that the gene was involved in producing the food stored in the egg. To understand eggs it will be necessary, at a minimum, to create mutations and carry out such studies on all 20,000 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 molecularly the gene under study becomes essential to make further progress. Cloning is simply a way of purifying an individual gene so that its DNA structure can be determined in the laboratory. If someone could invent a piece of filter paper that would allow only one specific gene to pass through while the other 20,000 stayed be-

hind, then gene cloning could be replaced by a coffeemaker.

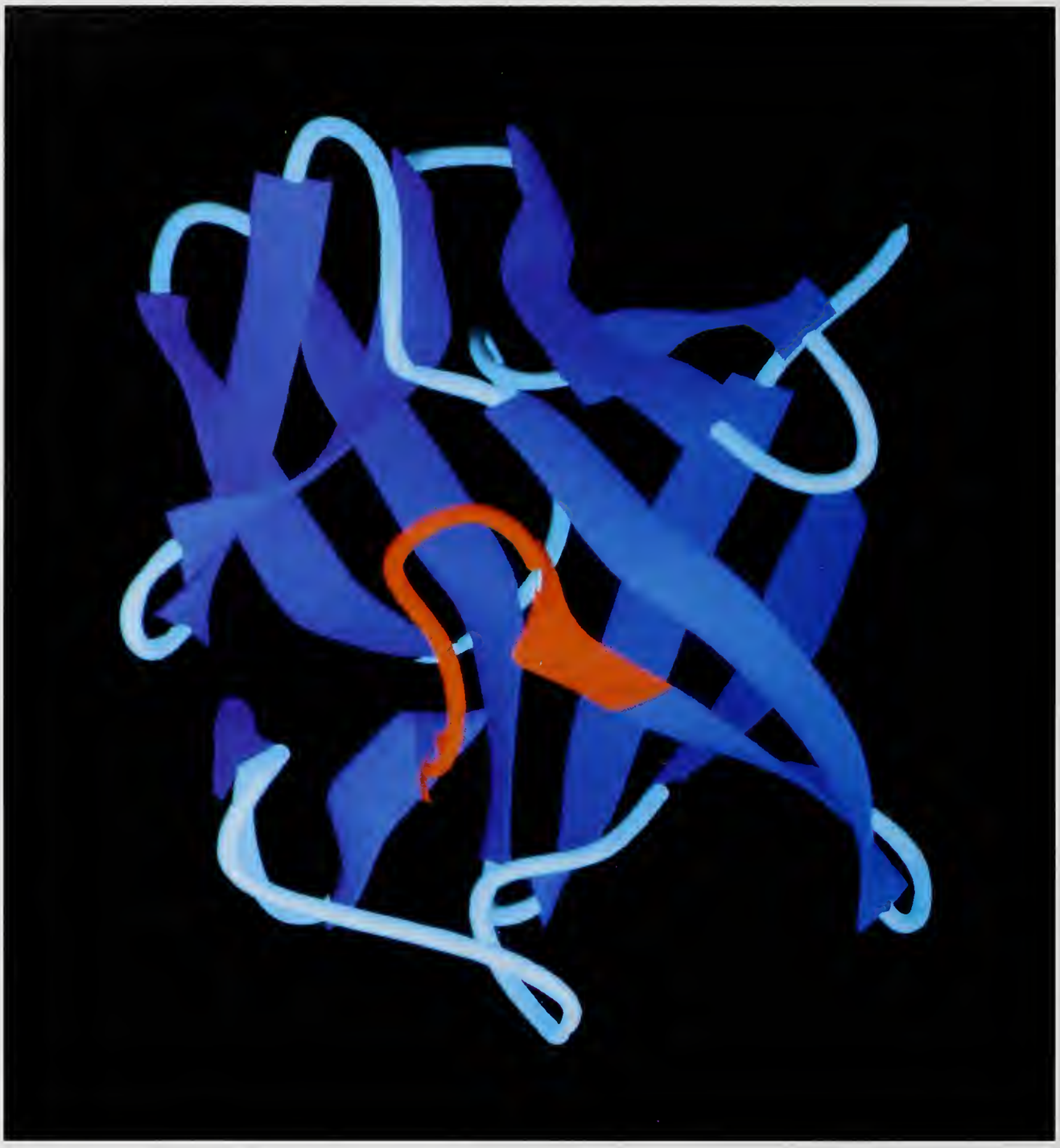
Gene purification has two major benefits. First, the structure of the gene's product, a specific protein that actually becomes part of the egg, can now be determined by DNA sequencing. Second, specific antibodies can be prepared that bind only to the product of that particular gene. When these antibodies are used, eggs that appeared normal in the microscope can suddenly be seen to have specific defects that are responsible for their inability to function normally. Such molecular biological studies allow us to understand in much greater detail where particular gene products are located in a egg, where they go as development begins, and what role they are likely to be playing in the process of development.

Covert Signals

One project in our research group concerns an unusual role played by the egg's nucleus during oogenesis. The nucleus of a cell contains its genes and is the site where gene activity is controlled. Normally, when a gene becomes active, it copies itself and releases the copies into the cytoplasm, where they instruct the cell how to make a particular protein. However, the nucleus doesn't usually target these copies to any particular region of the cell; they diffuse away from the nucleus in all directions through the cell cytoplasm. This year we have carried out several studies suggesting that *Drosophila* egg nuclei play a more sophisticated role during certain key times in oogenesis.

The oocyte nucleus always moves from a central location in the egg to a spot near the upper surface of the egg. To learn if this migration might be important, we destroyed the nucleus with a laser microbeam just before it moved upward, and allowed the eggs to continue developing. If the nucleus simply broadcasts products generally throughout the egg, its loss should cause only generalized defects, and possibly arrest further egg development. Instead, the upper surface of each treated egg was specifically affected; it now continued to develop as though it were the lower surface. (The two surfaces can be distinguished by differences in their curvature and in the eggshell.) This led us to suspect that the egg nucleus was moving to the upper surface in order to send specific information to that region. The most likely recipients of this information were some of the cells that surround the entire surface of the egg as it develops, called follicle cells. The egg nucleus appeared to be giving special instructions only to the follicle cells that lie just above the upper surface. Without these instructions these cells were unable to carry out functions that were necessary to produce a normal egg upper surface.

Once we suspected that the oocyte nucleus and the follicle cells were communicating, we sought to identify specific genes that might be involved in transferring developmental information. Several genes were known previously whose absence causes defects in egg development very similar to those produced by our laser treatments. For example, the gene *torpedo* encodes a type of receptor protein found in all the follicle cells that may receive the signal from the oocyte nucleus. We have subsequently identified, cloned, and begun to study a new gene, *zeppelin*, that is also required for upper surface development. This gene appears to be expressed preferentially in the upper follicle cells. It now appears feasible to identify most if not all the genes used by the oocyte to signal the upper follicle cells and to use them to learn exactly what the transmitted message consists of and how it is received and implemented.



Schematic of the three-dimensional structure of basic fibroblast growth factor, with the putative receptor binding site highlighted in orange.

From Zhang, J., Cousens, L.S., Barr, P.J., and Sprang, S.R. 1991. Proc Natl Acad Sci USA *88:3446–3450.*

Structural Studies of Proteins Involved in Hormonal Signal Transduction

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 one another by secreting growth factors, cytokines, or hormones and by responding to factors produced by other cells. These chemical messengers act by changing patterns of gene expression within the target cell, thereby altering its metabolic or developmental program. One of the major unsolved problems in biology is to define the mechanisms by which cells specifically recognize these chemical messengers and to understand how the messages are transduced intracellularly.

In our laboratory, we are using x-ray crystallographic techniques to define the molecular nature of the interactions between specific growth factors and their cognate receptors. These 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 receptors specifically recognize growth factors and how the receptor molecules' structure and chemical properties are changed, if at all, as a consequence of this interaction.

Tumor Necrosis Factors

One family of factors we study, called tumor necrosis factors (TNF), are produced by cells of the immune system. Macrophages—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 would occur in the event of infection. The avid attachment of this cytokine to receptors present on a variety of cells in the body triggers a series of events that dramatically alter fat and triglyceride metabolism and mobilize neutrophils. These events mediate inflammation, endotoxic shock, and the wasting phenomenon cachexia, from which the cytokine takes one of its names. TNF-α also induces a number of enzymes that effect the destruction and remodeling of tissues.

A related cytokine, TNF-β, is produced by T lymphocytes. It has many properties in common with TNF-α, including the ability to bind to the same receptors, but the two cytokines are only 30 percent identical in amino acid sequence. It is of interest to learn how such distantly related molecules can interact with the same receptor.

The three-dimensional structure of TNF-α, determined by graduate student Michael Eck, revealed the molecule to be composed of three identical protein subunits packed about a threefold axis of symmetry. We speculated that TNF may be capable of binding three receptors at once and that the TNF-induced aggregation of receptors might trigger the cascade of molecular events within the cell. The solution of the structure of TNF-β is still in progress, but initial results indicate that it has the same overall subunit tertiary structure and trimeric quaternary structure as TNF-α. Of particular interest, once more structural information is available about the nature of the TNF-receptor interactions, will be to determine whether different or similar chemical groups are used by the two cytokines in recognizing the TNF receptor. The proteins used in our crystallographic studies were provided by George Kuo of Chiron, Inc. (TNF-α), and Mark Ultsch, Bart DeVos, and Anthony Kossiakoff of Genentech, Inc. (TNF-β).

Recently several researchers have succeeded in isolating the gene encoding the receptors that engage TNF-α and -β. Two different types of TNF receptor were found. Surprisingly, their amino acid sequences bear little 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. The cysteine residues may form crosslinks to give each repeat a common and distinctive three-dimensional structure. Preliminary data from other laboratories suggest that each trimer of TNF-α interacts with three receptor molecules.

Collaborating with Bruce Beutler (HHMI, University of Texas Southwestern Medical Center at Dallas), we have initiated experiments to express the extracellular domains of TNF receptors in both mammalian and bacterial cells, which we will then crystallize in both the presence and absence of TNF-α and TNF-β. We expect that the three-dimensional structures will show how the

two receptors, though different in amino acid sequence, are able to recognize the same cytokines. We hope also to learn whether receptor-cytokine engagement induces significant changes in the receptor structure or the interactions among receptors that might mediate their internalization by cells.

Fibroblast Growth Factor

Basic fibroblast growth factor (bFGF), one of a group of seven structurally related proteins, triggers the induction of mesoderm formation in early embryogenesis. In adult animals, the same factor initiates processes of wound repair and vascular regeneration. It is present in the extracellular matrix, bound to heparin-containing proteoglycans, where it stimulates the movements of cells to sites of tissue generation or repair. The intracellular domain of the receptor for bFGF is a tyrosine kinase, an enzyme that initiates a cascade of intracellular reactions culminating in altered patterns of gene expression within the cell.

The three-dimensional structure of bFGF has been determined with high resolution by graduate student Jiandong Zhang, using protein samples provided by Lawrence Cousens of Chiron, Inc. We found that the molecule is structurally related to members of the interleukin-1 family that stimulate the growth of white blood cells. The structure reveals probable sites for receptor and heparin binding at separate but neighboring locations on the molecule's surface and also suggests how phosphorylation of bFGF might alter its affinity for the receptor.

We are collaborating with Phillip Barr of Chiron, Inc., and Lewis Williams (HHMI, University of California, San Francisco) to cocrystallize bFGF with its receptor's extracellular domain. At least five distinct receptors of bFGF have been identified, all of which bear extracellular domains related to the immunoglobulin family. We may find that the complexes between FGF and receptor dimers (which form when receptors interact with the factor) are structurally similar to those between antibodies and antigens.

G Proteins

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 membrane-bound enzyme that catalyzes the synthesis of the intracellular second messenger cyclic AMP. Gsα eventually hydrolyzes the GTP, allowing it to rebind its inhibitory subunits and return to an inactive state in which it can no longer stimulate cyclase.

We are undertaking crystallographic studies of Gsα to learn how this protein may interact with other components of the signal transduction system in which it operates. The initial steps toward a structural determination of Gsα have now been taken in our laboratory by research associate David Coleman, who has crystallized Gsα and a nonhydrolyzable GTP analogue in complex with a protein expressed in the laboratory of Alfred Gilman (University of Texas Southwestern Medical Center at Dallas). Our goal is to understand the series of GTP-dependent structural changes within Gsα that account for its regulatory properties.

The Production and Actions of Insulin and Other Islet Hormones

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 R. H. Williams at the University of Washington School of Medicine. After joining the faculty at 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 awards, including the Lilly and Gairdner Awards, and several honorary degrees.

THE production and regulated release of insulin is essential for normal growth and the utilization of food. A deficiency of insulin, or defects in its action, can lead to diabetes, a disease characterized by high blood sugar and a variety of possible complications, including blindness, heart disease, stroke, and increased susceptibility to certain infections. Diabetes affects 2–3 percent of the population of the developed countries.

Control of diabetes can be achieved through various combinations of diet, oral hypoglycemic agents, and/or insulin injections, depending on type and severity. Such therapies, however, are often less than fully satisfactory; they may only retard the progress of complications. A better understanding of insulin biosynthesis, its regulated release into the bloodstream, and its action on receptors to regulate metabolism and growth is vitally important to the development of new approaches to the therapy of this disease.

Insulin is made in the islets of Langerhans—small clusters of cells in the pancreas. Islet cells also secrete other hormones that appear to regulate digestion or nutrient metabolism. These include glucagon, somatostatin, amylin, and pancreatic polypeptide. The islet hormones, like many other regulatory peptides in the body, are derived from larger precursor molecules (proteins) called preprohormones. These precursors contain additional portions that serve as markers to guide them along special intracellular pathways, where they are concentrated into storage vesicles and processed into their smaller, biologically active forms. These are then secreted into the bloodstream in varying proportions to meet physiological needs. One goal of our research is to learn more about how prohormones like proinsulin are sorted from other proteins, concentrated into secretory granules, and processed into active hormones by specialized protein-splitting enzymes.

Insulin secretion from β-cells in response to elevated plasma glucose is a complex electrochemical process resembling the transmission of nerve impulses or the contraction 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. Upon depolarization of the membrane, these channels 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 sulfonylureas—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 to understand both their normal functions and their possible malfunction in some forms of diabetes.

We are also studying mutations that affect (pro)insulin structure. Some of these mutations, occurring in families with mild diabetes, give rise to abnormal insulin molecules with greatly reduced biological activity, as reflected in reduced binding to insulin receptors. Certain other mutations in the insulin gene primarily affect the conversion of proinsulin to insulin, leading to other familial disorders known as hyperproinsulinemias.

Insulin acts on tissues by binding to a large, complex receptor protein on the surface of cells. This activates a tyrosine kinase that alters many intracellular processes through a cascade of phosphorylations. Binding of insulin to the receptor also leads to its uptake and degradation in the liver and other tissues. This process, known as receptor-mediated endocytosis, may also play a role in the generation of some biological responses to insulin and in mediating the rapid removal of insulin from the circulation. By studying defective or modified precursor, hormone, and/or receptor molecules, we hope to learn more about normal islet hormone production and action and about derangements in these processes that can lead to diabetes or other diseases.

Our current areas of investigation include the following:

- mechanisms of intracellular sorting and proteolytic processing of hormone precursors,
- mechanisms regulating insulin secretion,
- mutant human insulin or insulin receptor genes and their functional significance,

- structure and function of the human insulin receptor,
- new hormones of the islets—e.g., the recently discovered amylin peptide, and
- evolution of insulin and insulin-like growth factors and their related receptors.

Identification of Prohormone-converting Enzymes

Using recently developed methods of specific DNA amplification, we have identified two cDNAs (PC2 and PC3) from neuroendocrine cells that encode proteolytic domains similar in structure to those of Kex2, a yeast prohormone-processing enzyme. We are attempting to learn more about the expression and function of these proteases in β-cells and to delineate their proteolytic properties. Our approach is to express the proteases in a variety of cells *in vitro*, along with prohormones such as proinsulin and/or POMC (proopiomelanocortin).

Studies on Potassium Ion Channels of β-Cells

Regulation of insulin secretion involves the coordinated control of ion channels in the β-cell membrane. We have recently isolated and characterized cDNA and genomic clones encoding a voltage-dependent K^+ channel isoform expressed in human islets and in a human insulin-producing tumor. This K^+ channel, designated hPCN1, consists of 613 amino acids and is related to the *Shaker* family of *Drosophila* K^+ channels. Two other human K^+ channel isoforms that we have isolated, hPCN2 and hPCN3, are homologous to hPCN1, with 55 and 65 percent amino acid sequence identity, respectively.

The electrophysiological characteristics of hPCN1 have been examined by voltage-clamp recordings of oocytes injected with synthetic hPCN1 RNA. These have revealed hPCN1 to be of the delayed rectifier type. The channel could function to restore the resting membrane potential of β-cells after depolarization and thereby contribute to the regulation of insulin secretion. Efforts are also under way to identify and characterize the ATP-dependent K^+ channel that plays a key role in initiating β-cell depolarization in response to glucose.

Studies of the Insulin Receptor

We have studied a mutation in the insulin receptor protein that gives rise to severe insulin-resistant diabetes. The defect in this patient is a single amino acid replacement that prevents the receptor precursor from being processed into mature receptor α- and β-subunits on the cell surface. Studies are now nearing completion on mutagenesis of the cleavage site of the human insulin proreceptor to determine the exact structural requirements for its processing by a specialized cellular protease to generate functional receptors. We are also using mutagenesis to evaluate more fully the functional consequences of disordered processing of the proreceptor.

We have developed methods for expressing human insulin receptors at very high levels in insect or mammalian cells in culture. Either normal or modified receptor proteins (generated by mutating the cDNA sequence) can be produced in amounts sufficient for detailed study of their structures, insulin-binding domains, and biological signaling properties. In other studies, we are analyzing the complexity of protein phosphotyrosine phosphatases—cellular enzymes that reverse insulin receptor activation.

Amylin, a New Product of the β-Cell

We recently identified a new β-cell precursor molecule that, when appropriately processed, gives rise to a peptide called islet amyloid polypeptide (IAPP), or amylin. This peptide 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 encoding IAPP and have begun studies on its biosynthesis and secretion in response to glucose and other factors. It is produced at a level about 1 percent that of insulin and is cosecreted with insulin. Studies on IAPP precursors from nine mammalian species support the hypothesis that differences in the sequence of IAPP in its central region (residues 20–29) in different species influences their relative susceptibility to the deposition of amyloid in the islets or in insulin-producing tumors (insulinomas).

The Evolution of Insulin and Insulin-like Growth Factors

We have succeeded in efforts to identify primitive insulin superfamily-related genes (encoding insulin or the closely related insulin-like growth factors IGF-I and -II) in primitive vertebrate organisms, including cyclostomes (the Atlantic hagfish) and amphioxus. The presence of a hybrid insulin/IGF molecule in amphioxus, a protochordate, suggests that the insulin-like growth factors diverged from an ancestral preproinsulin-like molecule in the very earliest stages of vertebrate evolution. We are also studying the origin and evolution of insulin and IGF receptors in these lower forms.

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, and the Warren Triennial Prize (shared with Thomas Cech). 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 roles in cells.

Systemic lupus erythematosus (SLE) is one of a number of diseases in which an individual's immune system mistakenly makes antibodies against the body's own molecules. Curiously, molecules that are very abundant in cells and highly conserved, 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 a number of different 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. For a gene's product to be made, the information in the gene's DNA is transcribed into an RNA copy (called pre-mRNA) that, after being "processed," directs the synthesis of a protein product. The DNA and the pre-mRNA contain segments called exons, which code for the gene's product, and segments called introns, which are noncoding regions interspersed between the exons. Before leaving the cell nucleus as mRNA, the pre-mRNA is cut, the exons are spliced back together, and the introns are discarded. The individual exons must be joined precisely and in the same order they originally occupied in the gene. Sometimes differences in the way exons are spliced can lead to different protein products in different 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 act to 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 together to align the exon ends so that precise splicing can occur. In this sense, snRNPs are much like ribosomal subunits (which also contain 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, particularly during the first step of intron removal. Here, we have investigated a bizarre type of splicing where exons from two separate gene transcripts are joined (trans-splicing) and have learned that trans-splicing does not require the U1 snRNP, because one of the transcripts contains intrinsic information that substitutes for U1. Investigation of the minimal sequences and/or structures needed will further elucidate the role of snRNPs in the spliceosome. The possibility that the most abundant nuclear RNP currently of unknown function, which contains 7SK RNA, may serve to recycle snRNPs after splicing is completed is likewise being actively pursued.

Mammalian cells also contain many minor

snRNPs that are closely related to the splicing snRNPs. One is the U7 snRNP, 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, which we anticipate all function in some aspect of mRNA metabolism, include viral snRNPs. For instance, in marmoset cells infected by *Herpesvirus saimiri*, which causes malignant transformation of the T cells of the immune system, there are five small RNAs encoded by the viral genome. They are hypothesized to contribute to either mRNA processing or stabilization, 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 ribosome assembly occur. Recently, in collaboration with Barbara Sollner-Webb's laboratory at the Johns Hopkins University School of Medicine, we established that the most abundant nucleolar snRNP (containing U3 RNA) is essential for the first step of ribosomal RNA processing. The possibility that other nucleolar snRNPs act in subsequent steps is now being investigated. A separate study is designed to dissect the signals that dictate the delivery of these 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 and 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. We have recently discovered a highly abundant, highly conserved cell protein that binds the EBERs and hope that its characterization will lead us to an elucidation of EBER function.

Finally, we are studying the enzyme telomerase, which synthesizes the ends of chromosomes of all higher organisms. In the unicellular ciliate *Tetrahymena*, telomerase has been shown to be an RNP. We have demonstrated telomerase activity in mammalian cell extracts and are now seeking to identify the mammalian snRNP. It appears to be responsible for the "healing" of broken chromosomes, as exemplified by certain thalassemias.

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 how they might be altered 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 Interaction

Thomas A. Steitz, Ph.D.—Investigator

Dr. Steitz is also Professor of Molecular Biophysics and Biochemistry and 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 Caltech. 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 62 other possible anticodons.

The protein domain that contains the active site has a structure similar to that of the analogous domains of the tyrosyl- and methionyl-tRNA synthetases and similar to another 7 of the 20 synthetases. Synthetases for 10 amino acids belong to a second unrelated class of synthetases.

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 a 30-bp DNA fragment and cAMP. The earlier CAP·cAMP structure had shown that each subunit of this dimer consists 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, this complex shows a severely bent DNA, with an overall bend of about 90°. Most of the bend is achieved by two large kinks of about 40° each. The relationships between this CAP-induced DNA bend and transcription activation are presently being pursued.

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 another structural domain, with a cleft large enough 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. The second crystal form, grown under conditions when the enzyme is active, is presumed to have DNA at the polymerase active site and has recently been solved. Although there are clear changes in the structure of the polymerase domain, no bound duplex DNA has yet been visualized. 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.

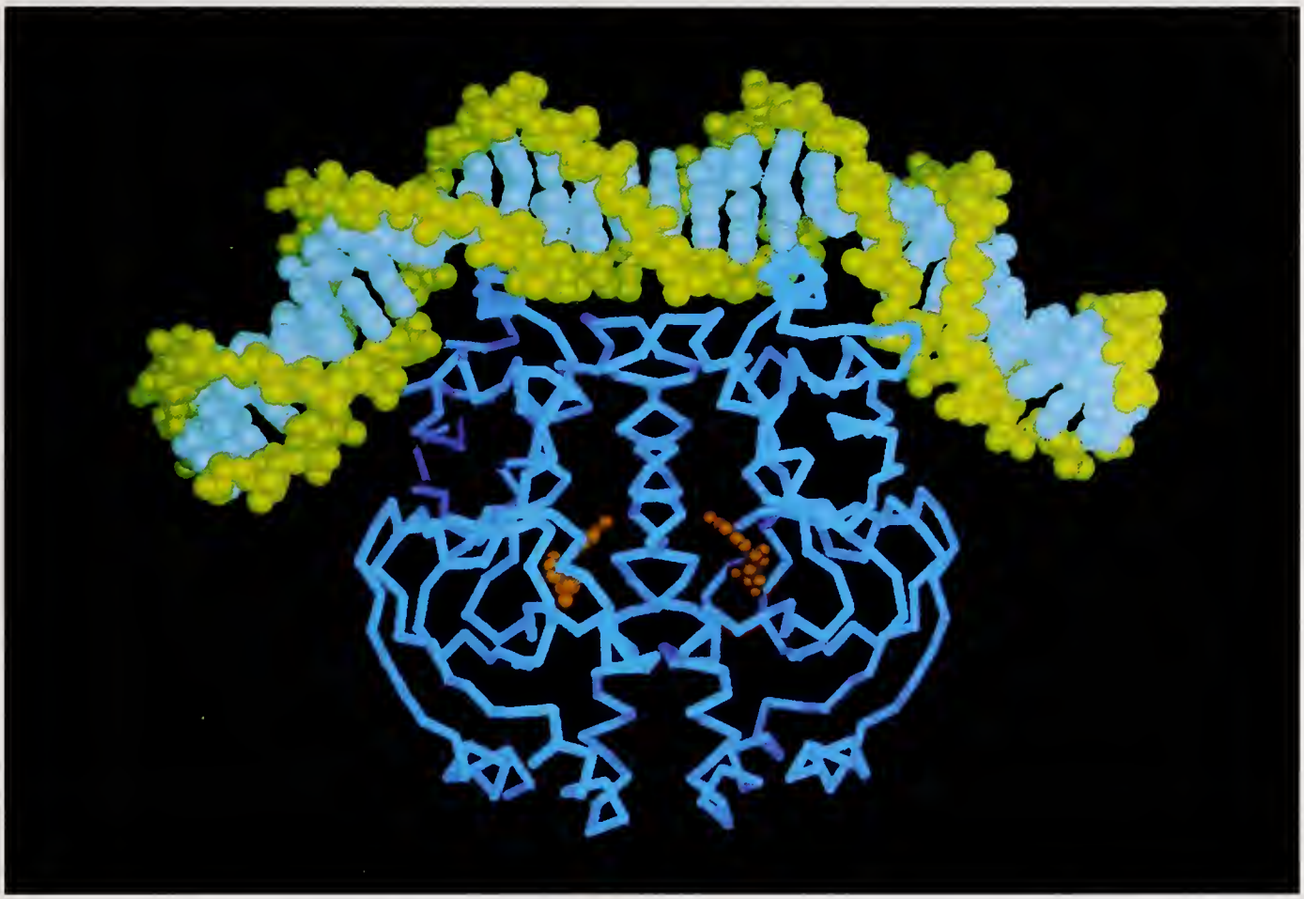
Genetic Recombination

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 but does not suggest a mechanism for this recombinase. We have recently co-crystallized this protein with a 31-bp fragment that contains the recombination site, whose structure should provide 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.

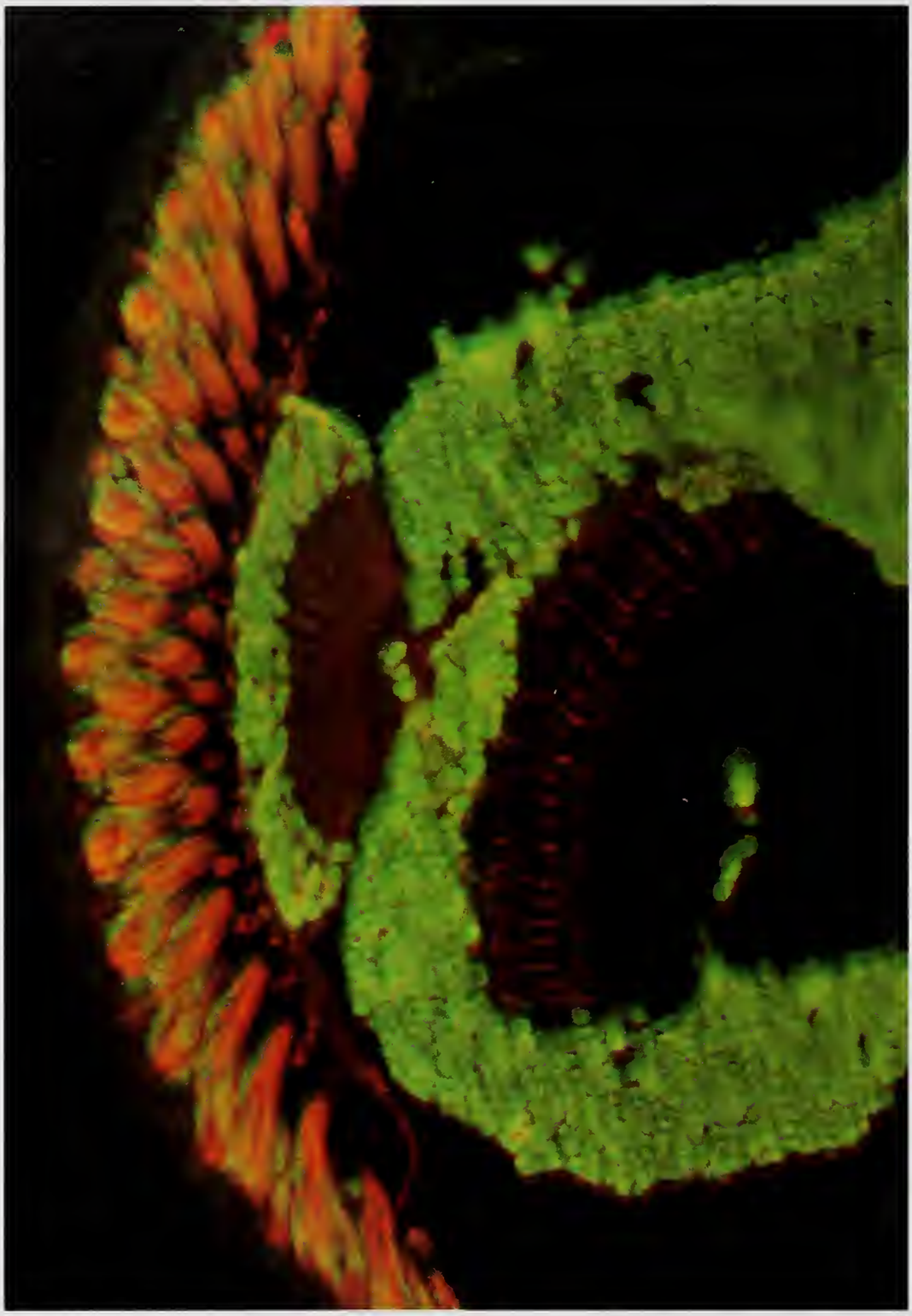
HIV Proteins

We have made large quantities of two proteins, reverse transcriptase and fragments of an activator protein called Tat, that are specified by HIV (human immunodeficiency virus). Toward our goal of crystallizing and establishing the structures of biologically relevant complexes, we have shown that fragments of Tat bind specifically to a portion of the HIV RNA called TAR. Initiation complexes formed using human $tRNA^{Lys}$, an HIV RNA fragment, and reverse transcriptase are being explored. We have recently calculated an excellent 5 Å resolution electron density map of reverse transcriptase and should soon extend this study to 3.5 Å resolution. Our long-term goal is to provide a structural basis for designing inhibitors of these proteins that might function as useful anti-AIDS drugs.



Representation of the crystal structure of Escherichia coli *CAP·DNA complex. The transcription factor, catabolite activator protein (CAP), is represented as an α-carbon backbone in blue. Bound cAMP is shown in red. The 30-bp DNA bound to CAP in the crystal is shown in partial space-filling representation with the bases in light blue and the sugar phosphate backbone in yellow. The DNA is bent 90° by CAP binding, initiated primarily at two 40° kinks.*

Research of Thomas A. Steitz.



Visualization of axonal projections from photoreceptor neurons of the compound eye to the optic ganglia in the brain of Drosophila. *This section has been stained with two different neuron-specific antibodies that recognize photoreceptor cell membranes (red) and neuronal nuclei (green), respectively. Photoreceptor axons can be seen to terminate in two distinct regions of the brain.*

From Selleck, S.B., and Steller, H. 1991. Neuron *6:83–99.*

Pattern Formation and Neuronal Cell Recognition in the Drosophila *Visual System*

Hermann Steller, Ph.D.—Assistant Investigator

Dr. Steller is also Assistant Professor of Neurobiology at the Massachusetts Institute of Technology and Adjunct Assistant Neurobiologist at the Massachusetts General Hospital. 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 Ruprechts-Karls University in Heidelberg. 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 goal of our research is to understand how the enormous diversity and specificity of individual neurons arise during development and how these neurons establish highly specific patterns of interconnections. We address these questions primarily by exploring the *Drosophila* visual system, where it is possible to apply powerful genetic and molecular techniques to study the development of individual cell types. We have identified mutations that affect the projection pattern of specific axons, the production of neurons from neuronal precursor cells, and the pattern of cell death, and have begun to characterize several of these at the cellular and molecular level. We hope that our results will eventually contribute to a better understanding of human neurological disorders.

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 has allowed us to screen systematically 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 identified a small number of loci that are required for the establishment of stable connections between the larval optic nerve and its target cells in the developing embryonic brain. One of the corresponding genes, *disconnected* (*disco*), has been analyzed at the molecular level and appears to encode a transcription factor. We are currently 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 into 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 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.

Innervation, 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 recently discovered 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, their continued survival requires these connections. This situation is reminiscent of trophic interactions widely described in vertebrates.

Molecular Genetics of Nematode Development and Behavior

Paul W. Sternberg, Ph.D.—Assistant Investigator

Dr. Sternberg is also Assistant 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 Searle Scholar.

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 tail. The concerted morphogenesis of male-specific neural cells, along with specialized muscle and epidermal cells, forms the copulatory male tail, which allows males to mate with hermaphrodites. Because *C. elegans* hermaphrodites are internally self-fertilizing—each animal producing both sperm and ova—male mating is a dispensable behavior. 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 more than 100 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.

Of the males with obvious abnormal tails, we have focused on those with abnormal spicules—innervated structures crucial to successful mating. Each of the two spicules comprises nine cells: two sensory neurons, one motoneuron, and six supporting cells. So far we have identified several genes that act at various stages in development of the spicules. For example, two genes are necessary for production of the hardened cuticle that surrounds each spicule.

We have identified several genes that are required for the decision between alternative neuroblast fates. Two cells, Bα and Bβ, each normally generate distinct sets of spicule cells. Two genes promote differentiation as Bβ, while four others promote differentiation as Bα. Of the former class, the *let-60* gene (lethal gene 60) is of special interest because we have discovered that the level of its activity determines whether the neuroblasts in question become Bα or Bβ: low *let-60* activity makes a cell Bβ, high *let-60* activity makes a cell Bα. This set of genes also controls the fates of two other sets of cells whose fates during development depend on signals from neighboring cells, suggesting that these genes encode a general mechanism that controls the fates of cells during development. The products of *let-23* and *let-60* are homologous to mammalian proto-oncogenes, epidermal growth factor (EGF) receptor and *ras*, respectively, suggesting that common mechanisms might control the fates of cells in mammals and nematodes.

Male mating is the most complex nematode behavior. By studying the Cod mutants, we hope to elucidate how genes control each step in this 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 24 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 3 will endlessly circle the hermaphrodite searching for the vulva. A minority of the mutants are ineffective at more than one step. Having mutants blocked at defined steps will allow us to identify genes and cells necessary to specify this innate behavior.

We are also identifying the cells responsible for each step in mating behavior by killing individual cells or structures with laser microbeam irradiation and observing the consequences for behavior. For example, the spicule motoneuron and at

least one of the two spicule sensory neurons are necessary for spicule insertion.

The establishment of cellular asymmetry is a fundamental yet relatively unexplored aspect of cell regulation. We have begun to study this problem in the context of the male B lineage, which generates the spicules as described above, and in a cell called the 2° vulval precursor cell. The *lin-18* gene (cell lineage gene 18) is necessary for the asymmetry of the 2° vulval precursor cell and is the starting point for molecular genetic studies of this problem. 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 well-characterized signal that instructs these cells to be 2° cells, and not other types of cells.

In a project initiated this year, we are analyzing the function of the nematode homologues of signal transduction molecules identified in mammals. We have cloned the nematode gene similar to the *raf-1* protein kinase and placed it on the *C. elegans* genomic map. To identify the cells expressing G protein genes, we have constructed chimeric genes consisting of the transcriptional control regions of two G protein α-subunit genes linked to a "reporter" gene. We then create transgenic nematodes by introducing the engineered genes back into the nematode through microinjection of DNA. Both genes are expressed primarily in neurons, including neurons in the male tail. We plan to use these chimeric genes as markers for neuronal types to examine the molecular anatomy of the Cod mutants.

How the Brain Modifies Its Own Circuits: Synaptic Neuromodulation

Charles F. Stevens, M.D., Ph.D.—Investigator

Dr. Stevens is also Professor of Molecular Neurobiology at the Salk Institute and Adjunct Professor of Pharmacology at the University of California, 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.

A remarkable feature of the brain is its ability to modify itself. Neuronal activity can change the properties of nerve cells so that the neural circuits responsible for the brain's computations become physically different. This ability is termed *neuromodulation*.

To understand neuromodulation one must have an idea of how neuronal circuits are constructed and how signals are passed from one brain cell (neuron) to another in these circuits. Each neuron in the brain receives information from about 10,000 neurons and, in turn, sends information to about 10,000 others. Information is transmitted from one nerve cell to the next at special points of contact known as *synapses*. Neural circuits, then, are formed by chains of neurons that are interconnected at synaptic contacts in incredibly complicated ways.

Each time a nerve impulse arrives at a synapse, the sending cell releases a small quantity of a chemical known as a *neurotransmitter*. The neurotransmitter diffuses rapidly—in less than a thousandth of a second—to the receiving neuron, where it binds onto special proteins, called *receptors*, embedded in the receiving cell's surface membrane. When the neurotransmitter binds onto the receptor, it causes it to change shape and to reveal a submicroscopic pore through which an electric current can flow. This current, carried by ions, constitutes the signal within the receiving neuron and is used to determine when the receiving cell itself will produce impulses.

Two large classes of receptors are known: excitatory and inhibitory. In the receiving neuron, excitatory receptors cause changes that tend to produce nerve impulses, and the inhibitory receptors do the opposite: they prevent the generation of nerve impulses. Thus the two classes of synapses act as an accelerator and a brake on nerve cell activity. About two-thirds of the synapses in the brain are excitatory, and the rest are inhibitory.

An important determinant of how the circuits function is the strength of the various synaptic contacts. If a synapse is strong, it will be effective in helping to produce (or prevent, if inhibitory) nerve impulses in the receiving cell, whereas weak synapses will be much less effective. Clearly, what a neuronal circuit computes will depend on the strengths of its various synapses.

Our laboratory has discovered that the strength of excitatory synapses is under control of the brain activity itself. Thus the efficiency with which one neuron transmits information to another is subject to neuromodulatory control of still other neurons. This control is quite specific and can be targeted to particular neurons in a circuit.

The mechanism of this neuromodulatory control is a version of one very common in biochemistry. A special enzyme known as protein kinase A (PKA) adds a phosphate group to the receptor, and this modified receptor then has a greater effect on the receiving neuron. To understand this mechanism more completely requires some additional information about how receptors work.

When the receptor binds its neurotransmitter —glutamate, in the case of the excitatory receptors—and opens its pore for ions to flow through, the receptor usually remains in its open state for about a thousandth of a second. Techniques are available that permit the neurobiologist to record the currents that flow through single receptors so that we can observe them as they open and close.

When PKA is applied to small patches of membrane containing excitatory receptors, these receptors are modified: the altered (phosphorylated) receptors dwell in the open state 5 to 10 times longer than usual, and the signal received at the synapse is 5 to 10 times larger for each impulse arrival. The increased strength thus occurs through increases in the open time of the receptors.

But what controls the PKA? Our laboratory has found some, but not all, of the answers to this question. PKA is activated by a rather complicated cascade of events that involve the familiar messenger molecule cAMP. This cascade is ultimately engaged by the action of chemicals known as *neuromodulators* that are released by other neurons when they are active. Many neuromodu-

lators operate through the cAMP cascade, but the specific neuromodulator used in this instance, and the particular neuronal pathway employed for the neuromodulatory signal, have not yet been identified.

The experiments described above used neurons from a particular region of the brain—the hippocampus—that were grown in tissue culture. In order to be sure that the neuromodulatory mechanisms investigated were not some peculiarity of cultured neurons, we have also demonstrated that the same effects can be found in brain slices of the hippocampal region.

Thus our laboratory has discovered a powerful mechanism through which neuronal activity can regulate synaptic strength and thereby modify the computations carried out by neural circuits. The next challenge is to discover how the brain actually makes use of this mechanism to switch from one compensational state to another.

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 P. A. Lawrence at Cambridge. Before joining the faculty at Columbia, he conducted molecular and genetic studies in Drosophila *with T. Maniatis at Harvard University. In addition to several fellowships, he received the McKnight Neuroscience Development Award.*

MOST animals large enough for us to see with the naked eye consist of billions or trillions of individual cells arranged in extremely complex and highly conserved patterns. To generate such patterns, cells in a developing organism must know where they are and choose what to do based on this information. Research in our laboratory is directed toward understanding how such spatial information is generated and interpreted during development.

For many years considerable attention has been given to the concept that gradients of diffusible "form-generating" substances, or morphogens, might provide this spatial information. For example, a localized source of a diffusible, but unstable, morphogen would generate a stable gradient, which could, in turn, provide a series of concentration thresholds that would determine distinct cellular behaviors (e.g., the development of particular pattern elements) as a function of distance from the source.

During the past 10–15 years, the pioneering embryologic and genetic experiments of Christiane Nüsslein-Volhard and her colleagues have established that several distinct morphogens specify the basic body plan of the fruit fly *Drosophila* early in development. One of these, the protein product of the gene *bicoid* (*bcd*), controls most aspects of the anterior pattern of the body, while the products of two other genes, *nanos* (*nos*) and *torso* (*tor*), seem to specify the pattern of the remaining posterior and terminal portions. Our principle aim is to determine the roles and modes of action of each of these different morphogen systems.

Control of Anterior Body Pattern by the *bcd* Morphogen

Dr. Nüsslein-Volhard and her colleagues established that transcripts of the *bcd* gene are synthesized and then transported to the anterior pole of the egg during oogenesis. Following fertilization, these tightly localized transcripts are translated, and the resulting protein diffuses from its site of synthesis, generating a concentration gradient. To determine how the *bcd* protein gradient dictates anterior body pattern, we have sought to identify direct targets of *bcd* action and to determine the mechanisms that allow different concentrations of protein to control how these targets respond.

In an initial series of experiments, we established that the *bcd* gradient is responsible for generating an opposing gradient of another regulatory protein, *caudal* (*cad*). Although we have yet to determine the exact mechanism by which this response occurs, our data suggest that *bcd* protein may repress translation of maternally derived *cad* transcripts that are ubiquitously distributed in the early embryo. More recently, we have identified a second target of *bcd* action, and in this case determined the mechanism involved.

The target is a zygotic regulatory gene, *hunchback* (*hb*), which is normally activated in a well-defined anterior portion of the early embryo. We have shown that *bcd* protein directly binds and transcriptionally activates the *hb* gene in a fashion that depends critically on *bcd* protein concentration. Hence, the anteroposterior gradient of *bcd* protein provides a distinct threshold dictating where the *hb* gene is expressed.

These and other experiments have also provided evidence that the *bcd* gradient triggers a series of other distinct responses by the same mechanism—each response governed by the particular binding affinity of a different target gene for the *bcd* protein. The ability of the *bcd* protein to control the patterns of subordinate signaling molecules by at least two distinct mechanisms, as well as the demonstration that it acts directly on several distinct targets, clearly indicates how the graded distribution of a single morphogen can provide a complex set of instructions for organizing body pattern.

Control of Posterior Body Pattern by the *nos* and *hb* Morphogens

The development of posterior pattern is largely dependent on the *nos* morphogen, which appears to spread from a tightly localized source at the posterior pole during early embryogenesis. It was initially thought that this factor operates like

the *bcd* morphogen in directly activating a number of subordinate signaling molecules in specific domains of the early embryo. However, we have recently conducted experiments showing that this view is incorrect and have provided evidence that *nos* functions only to ensure the graded expression of another morphogen—surprisingly, the *hb* protein.

As discussed above, graded *bcd* protein directly activates transcription of the *hb* gene in a broad anterior domain of the embryo. The *hb* gene, however, is also transcribed maternally: upon fertilization, maternal *hb* transcripts situated in the posterior half of the embryo appear to be blocked translationally and subsequently destabilized by localized *nos* activity. Thus, like *bcd*, *nos* also generates differential *hb* protein expression along the anteroposterior axis, albeit by a different molecular mechanism.

Differential *hb* activity, whether generated under the control of the *bcd* or *nos* determinants, appears to be involved in specifying most of the posterior body pattern. To test whether *hb* protein functions as the primary gradient morphogen organizing the posterior body pattern, a series of mutant embryos have been generated that lack both the anterior (*bcd*) and terminal (*tor*) morphogen systems, but retain the posterior system (*nos*) and maternal *hb* transcripts derived from 0 to 4 copies of the *hb* gene. Under these circumstances, the formation of posterior body pattern is solely dependent on the graded distribution of *hb* protein derived from maternal transcripts. Moreover, different threshold concentrations of *hb* protein can be shown to determine distinct boundaries of expression of subordinate regulatory molecules, most notably the gap gene proteins *Krüppel* (*Kr*), *knirps* (*kni*), and *giant* (*gt*), which specify the position and polarity of most of the abdominal segments.

These experiments indicate that posterior body pattern is established under the control of an *hb* morphogen gradient, much as anterior body pattern is controlled by the *bcd* gradient. Since the *bcd* gradient can itself create an *hb* protein gradient by its direct and localized effect on *hb* transcription, it has the capacity in principle to specify virtually the entire body pattern. These findings reveal a basic asymmetry in the control of the anteroposterior body plan and provide an indication of how a relatively simple morphogen gradient like *bcd* can generate a more elaborate and long-ranging system of spatial cues.

Control of Terminal Body Pattern by the *tor* Receptor

Recent studies in other laboratories have established the existence of a third primary determinant system that distinguishes end portions of the body pattern from the middle. A critical determinant in this system appears to be a receptor tyrosine kinase encoded by the *tor* gene. In studies performed in this laboratory, the *tor* protein has been shown to be expressed uniformly along the surface of early embryos, even though genetic experiments indicate that it must normally be active at both poles but inactive in the middle portion of the body. Hence the *tor* kinase, as anticipated from its deduced molecular structure, appears to function as a ubiquitous surface receptor that is activated only in the vicinity of the poles.

Additional genetic and immunohistochemical experiments have confirmed this conclusion and suggested that *tor* is activated by a localized ligand initially tethered to the extracellular matrix surrounding the embryo. Finally, using a mutation of the *tor* gene that causes temperature-sensitive constitutive activity of the receptor, different levels of *tor* activity have been shown to specify distinct portions of the terminal pattern. Thus polarized activity of the *tor* kinase at the surface of the embryo may specify terminal pattern by generating one or more gradients of intracellular signaling molecules—perhaps phosphorylated substrates of the protein. The specification of terminal pattern may therefore differ from that of anterior and posterior pattern only in the mechanism by which graded expression of the active morphogen is generated.

Molecular Mechanisms of Neurotransmitter Release

Thomas C. Südhof, M.D.—Associate Investigator

Dr. Südhof is also Associate Professor of Molecular Genetics at the University of Texas Southwestern Medical Center at Dallas. He received his medical degree from the Georgia Augusta University of Göttingen, FRG. He obtained postdoctoral training first with Victor Whittaker at the Max-Planck-Institut für biophysikalische Chemie in Göttingen and then with Michael Brown and Joseph Goldstein in Dallas.

MANY cells send out signals by secreting messenger molecules. These molecules are recognized by specific receptors on target cells. The most striking example of this communication process is found in neurotransmission—the relay of information from one neuron to another in the nervous system. Neurotransmission occurs at synapses, highly specialized contact sites between two nerve cells. At a synapse, a presynaptic neuron secretes a signal that is recognized by the adjacent postsynaptic cell. A similar communication process is also responsible for the regulation of body metabolism by hormones, except that in this case several messenger-secreting endocrine cells are clustered together and send their signals to widely distributed cells in the body.

My laboratory explores one aspect of this important communication process—the mechanism by which signal molecules like neurotransmitters or hormones are released from cells. These molecules are synthesized in the secreting cells and packaged into intracellular membrane-bound vesicles. Upon stimulation, the cell's signal molecules are released by exocytosis, which consists of the fusion of the vesicle membranes with the plasma membrane and the concurrent secretion of the vesicle contents. Our goal is to characterize the components and mechanisms of the cell that are involved in the assembly, exocytosis, and recycling of synaptic vesicles.

The first goal in exploring synaptic vesicle functions consists of the characterization of their protein constituents. The cellular processes leading to the accumulation of neurotransmitters in synaptic vesicles and their release by exocytosis must act on the synaptic vesicle membrane. Therefore we are focusing on the proteins of synaptic vesicles, with the aim of understanding their structure and function in the context of membrane traffic in neurons.

Since synaptic vesicles have rather specialized functions, they are comparatively simple organelles. Preliminary estimates suggest that synaptic vesicles have fewer than 50 major protein components. Although the composition of these organelles has not yet been determined in molecular detail, they may be amenable to such an analysis. In addition to providing insight into synaptic functions, the study of the molecular structure of these vesicles will provide insight into the structures of organelles in general.

The membrane proteins of synaptic vesicles fall into two basic functional categories: transport proteins instrumental in the active uptake of neurotransmitters by synaptic vesicles, and structural proteins that mediate the vesicles' interactions with other cellular membranes and the cytoskeleton during exocytosis and endocytosis of the vesicles. The number and abundance of proteins in the second class is much greater than in the first class. This correlates with the greater complexity of the transport, fusion, and fission reactions of synaptic vesicles as compared with their neurotransmitter uptake processes.

In the past several years, more than 10 proteins that are abundant in the synaptic vesicle membrane have been characterized in my laboratory. The transport functions of synaptic vesicles depend on a proton pump in their membrane; the activity of this pump establishes an energy gradient across the vesicle membrane. This energy gradient is utilized for the active accumulation of neurotransmitters. Two subunits of the synaptic vesicle proton pump have been cloned in my laboratory, and we are currently studying its assembly and regulation. Some synaptic vesicles also contain an electron transport protein in their membrane to convey reducing equivalents for intravesicular enzymes. This protein, cytochrome b561, has also been purified and cloned in this laboratory. We are now in the process of trying to identify and characterize the actual neurotransmitter transport proteins in the synaptic vesicle membrane.

The second class of proteins are those presumably involved in the exocytosis and re-endocytosis of synaptic vesicles. These proteins should be present on all vesicles independent of their neurotransmitter content. The most abundant among these are the three integral membrane proteins—synaptotagmin (M_r 65,000), synaptophysin (M_r 38,000), and synaptobrevin (M_r 18,000)—and the synapsins, a group of four related peripheral

membrane proteins. The primary structures of these proteins were determined in this laboratory, and specific antibodies and different expression systems are being used to probe their functional roles.

Synaptophysin has a complex structure consisting of a homotrimer that spontaneously re-forms into a homohexamer upon reduction of its disulfide bonds. Synaptotagmin and synaptobrevin are highly conserved synaptic vesicle proteins that we have also characterized in the fruit fly *Drosophila melanogaster*. Synaptotagmin contains two copies of a Ca^{2+}-binding domain found in protein kinase C. Recombinant synaptotagmin also avidly binds phospholipids, suggesting that it may have a role in the Ca^{2+}-regulated steps of synaptic vesicle exocytosis.

The synapsins are the major phosphoproteins of the nerve terminal and interact with the cytoskeleton in a phosphorylation-dependent manner. The four synapsins are the differentially spliced products of two separate genes. They are composed of mosaics of shared and unique domains and are differentially distributed among synapses. The synapsins presumably have a role in the vesicle traffic to and from the plasma membrane and may regulate the number of vesicles available for exocytosis. They are stoichiometrically phosphorylated at multiple sites by multiple protein kinases, as a function of the activation state of the nerve terminal.

Together, our studies have characterized a sizable portion of the synaptic vesicle proteins and are approaching the dissection of the synaptic vesicle pathway in molecular terms. Our current investigations of the interactions of synaptic vesicle proteins with other proteins of the synaptic nerve terminal are an attempt to understand the role of these proteins in regulating neurotransmitter release.

Genes Responsive to Growth Factors

Vikas P. Sukhatme, M.D., Ph.D.—Assistant Investigator

Dr. Sukhatme is also Associate Professor of Medicine and 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.

THE primary focus of our research is to understand the molecular events involved in the regulation of mammalian cell growth. Since cells grow in response to the presence of extracellular mitogens (growth factors), we are attempting to define the mechanisms by which growth factors function. A complex series of events follows mitogen-receptor binding, including the rapid generation of second messenger signals in the cell's cytosol and plasma membrane. In turn, these events lead to the activation of a set of "immediate-early" genes.

During the past few years, work from our own and other laboratories has shown that an important subset of these genes encodes transcriptional factors—proteins that activate or repress another set of genes. Furthermore, upon mitogenic stimulation these genes are inducible in a wide variety of mammalian cells, suggesting that they are part of a general program involved in cell growth. More importantly, by virtue of their structure and induction kinetics, they are likely to play broad roles as "third messengers," by coupling early biochemical processes to long-term changes in gene expression required to modulate not only cell growth but other cellular processes such as differentiation. Therefore, our long-term goal is to understand how these proteins function as nuclear signal transducers.

To date, our laboratory has identified four early growth response (EGR) genes (EGR1–4). Each encodes a protein with three Cys_2-His_2 zinc finger motifs, structures previously described in other transcriptional regulatory proteins. The predicted finger domains are nearly identical to each other and are closely related to that of another protein that may play a critical role in the genesis of a childhood kidney tumor (Wilms' tumor). These proteins function as transcription factors and bind to similar target sequences. Thus they can interfere with each other's ability to activate a test target gene. However, the physiologically important target gene(s) for the EGR proteins in the context of cell growth remains to be identified.

We are currently attempting to define a functional role for the EGR1 transcription factor in myeloid differentiation and in tumorigenesis. The human EGR1 gene located on chromosome 5 band q31 is deleted in the blast cells of many patients with therapy-related acute myeloid leukemia. These patients succumb to this disease after successful treatment of their primary malignancy years earlier. Recent fine-mapping data continue to support the hypothesis that EGR1 is located in the critical region of chromosome 5q31, the smallest common deleted region in these patients. If these data continue to hold up as the critical region is further refined, we will use pulse-field electrophoresis and blotting as well as polymerase chain reaction methodology and DNA sequencing to investigate the structure of the cytogenetically normal EGR1 allele. Toward this end, we have recently cloned and sequenced the human EGR1 cDNA. The overall hypothesis under consideration is that, like the retinoblastoma gene, the p53 gene, and the Wilms' tumor gene, EGR1 might be a negative regulator of cell growth, at least in the context of myelopoiesis.

Another emerging focus of our laboratory is the isolation of genes for transcription factors involved in the development of a solid organ, the kidney. Using oligonucleotide probes derived from conserved regions of zinc finger proteins, we have identified several clones expressed at high levels in the kidney. Molecular characterization of these clones and their expression patterns during development are in progress. As an alternative approach, we are also attempting to characterize the genetic elements responsible for expression of the Tamm-Horsfall protein. This glycoprotein is produced in all mammalian species, in a developmentally regulated and nephron-specific manner, and should serve as a useful starting point for defining aspects of kidney transcription and development.

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 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 organ 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.

Early studies identified several genes that are rapidly expressed in response to growth factors in liver cells and in commonly studied cells like lymphocytes and fibroblasts, respectively important in the immune system and scar formation. 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.

To isolate novel genes involved in liver cell growth, we have used a technique that allows subtraction of genes expressed in nongrowing cells from those expressed in growth-stimulated cells. Thus genes that are uniquely expressed in growing liver cells have been isolated and can be studied. We have identified approximately 50 such genes through analysis of their sequence and through predictions concerning their protein structure. Many of the genes are expressed in other growing cells and seem to be part of the general growth response, while some are specific to liver.

In the past year, 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 are 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 alone among growing cells, and we found several in this category, including a few that are highly expressed. These proteins could be important in liver-specific growth.

The fourth category includes several genes whose expression is induced in regenerating liver but abnormally high in the liver tumor cell line that grows in response to insulin, even prior to insulin treatment. These could function as oncogenes in the liver tumor line.

LRF-1 (liver regeneration factor) is one of the proteins that we have further characterized. It 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, or with related proteins. LRF-1 has a characteristic temporal expression pattern during liver regeneration and 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.

We have begun studying the function of a protein secreted during liver regeneration that could have a role as a growth factor. We found that a gene encoding a protein that binds to so-called insulin-like growth factor, termed insulin-like growth factor–binding protein-1 (IGFBP-1), is highly expressed during the early phases of liver regeneration. IGFBP-1 is not expressed in other growing cells we have examined. This protein has been implicated in enhancing the growth-promoting effects of insulin-like growth factor. Because IGFBP-1 could potentially act as a growth factor during liver regeneration, we have further studied its expression and function in liver regeneration.

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. After doing 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 held laboratory positions 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 types of cells: 1) the B cell, which secretes antibodies that bind to and target foreign substances for elimination by the immune system, and 2) the T cell, which is capable of producing a variety of molecules, known as lymphokines, that regulate the functions of other cells involved in immune responses. T cells are also capable of identifying cells expressing foreign proteins and destroying them by cell-mediated cytotoxicity. 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.

The central role of the lymphoid immune system in the natural resistance to infectious diseases has been demonstrated by the infectious problems encountered by patients with both congenital and acquired immunodeficiency syndromes. Many of the serious infectious and neoplastic complications associated with the acquired immune deficiency syndrome (AIDS) are the result of depletion of the helper cells, a subset of 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.

B cells derive their name from the bursa of Fabricius, a developmental organ in birds that is required for the generation of mature B cells. Mammals lack a bursa of Fabricius but still appear to be able to generate a B cell immune system. Over the past several years, our laboratory has investigated the role of this organ in B cell development in the chicken. We are attempting to characterize differences in the generation of B cells in mammals and avian species to account for this lack in mammals. 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.

Our studies have helped demonstrate that the primary molecular mechanism by which chickens generate this large number of different antibody molecules is different from that used by mammals. Antibody diversity in the chicken is achieved by gene conversion, a molecular mechanism that occurs during B cell development in the bursa of Fabricius and that requires the induction of cell proliferation within the bursa of Fabricius. We have begun to characterize the requirements for this proliferation. These studies have led to insights into how genes encoding immunoglobulin molecules are initially assembled and modified during development.

It has recently been shown that immunoglobulin gene conversion also occurs in mammals. The molecular events that are used to generate gene conversions in chickens may also be used to create the somatic mutations that arise in mammalian immunoglobulin genes during B cell development. Somatic mutations have been implicated in enhancing the ability of a cell to respond to foreign substances but also in the development of leukemia and lymphomas. Our studies in the chicken should teach us more about this important but potentially dangerous process.

T cells have been divided in the past 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. A number of years ago the molecular requirements for the growth and expansion of cytotoxic T cells were defined. Those studies led to the discovery of the lymphokine interleukin-2 (IL-2) and, subsequently, to a variety of therapeutic trials using IL-2 to increase cell-mediated immunity in patients with cancer. Unfortunately, conditions for the growth of helper T cells have yet to be defined. An understanding of the molecular mechanisms that regulate the function and proliferation of helper T cells is now of particular interest, since it is helper T cells that are deficient in patients with AIDS.

Over the past several years, we have been investigating molecular mechanisms that regulate helper T cells. We have defined a helper T cell activation pathway that appears to regulate the

coordinate production of a variety of lymphokines. Activation of this pathway involves CD28, a specific receptor on a T cell's surface. The molecule that normally stimulates this activation pathway is expressed on B cells, macrophages, and thymic epithelial cells. Thus this receptor-ligand interaction serves as a means for different cell types involved in the immune response to communicate with each other.

Within the T cell, the CD28 activation pathway appears to control the rate at which the mRNAs for these genes are degraded, rather than the rate at which the mRNAs are produced. Stimulation of this pathway after mitogenic activation of cells produces exponential growth of helper T cells. As a result, we have been able to keep helper T cells in exponential growth for over six months. This is exciting, because for the first time we are able to grow normal human helper T cells in culture for long periods of time. In theory, we can now generate an infinite number of helper cells. Cells grown in this way are being tested for their ability to enhance an individual's immune response to cancer. In addition, we may be able to develop therapies to enhance the production of helper T cells within the body. Presently, however, we do not know whether the activation pathway directly leads to the proliferation of helper T cells or whether a particular lymphokine or combination of lymphokines produced as a result of this stimulation is responsible. Studies to clarify this issue are under way.

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 Medical Center. He obtained his undergraduate degree in biology from Colgate University. He received his Ph.D. degree in biochemistry at the University of California, Berkeley, where he worked with Robert Tjian. 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 higher eukaryotes. Three-quarters of a century of biological, physiological, and genetic experiments, combined with recent intensive molecular studies, have 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 a 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, which 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 being actively transcribed are often represented by large regions of decondensed chromatin that can be seen as puffs in these chromosomes. Thus the transcriptional activity of specific genes at specific times can be followed 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 larval development, when the burst of ecdysone triggers a dramatic change in the puffing pattern coincident with the onset of metamorphosis. 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 more than 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, as well as addressing the question of how gene hierarchies are controlled during development.

Our current studies center around *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 amino-terminal domains are rich in acidic amino acids, whereas the common carboxyl-terminal domain is rich in basic amino acids. This charge distribution is reminiscent of yeast transcriptional activators and thus consistent with the notion that *E74* may encode regulatory proteins. In addition, 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 oncogene-related proteins, such as *E74*, may help us learn more about how the normal counterparts of these disease genes function during development.

Biochemical analysis of the *E74A* protein has revealed that it binds DNA in a site-specific man-

ner and that its target sequence resembles that recognized by other ETS-domain proteins. By using antibody detection techniques to localize the *E74A* protein bound to the giant polytene chromosomes, we are now moving to the next step in the regulatory hierarchy. A variety of binding sites have been identified in this manner, including both early and late ecdysone-induced puffs. These target genes will be used for biochemical studies to determine the function of *E74A* DNA binding. In addition, we plan to characterize some target genes to learn what role they might play in metamorphosis.

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 does not appear in the cytoplasm, however, until 1 hour after promoter activation. This delay corresponds 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. *E74A* and *E74B* also show different patterns of transcription in four ecdysone target tissues. 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.

E74 provides a paradigm for our studies on the molecular basis of metamorphosis. We hope to use this gene to begin to understand how the hormonal signal is transduced to result in specific morphogenetic changes. Our next step is to determine which genes are regulated by *E74* and to characterize the products they encode, as well as their mode of regulation by *E74*. Our long-term goal is to isolate more early ecdysone-inducible genes to extend our understanding of this complex developmental process.

The Regulation of 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 the 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 at the NIH was done with Philip Leder. Before joining the faculty at Princeton, Dr. Tilghman held positions at Temple University and the Institute for Cancer Research, Philadelphia.

THE orderly development of the mammalian embryo requires the appropriate activation and subsequent modulation of genes in a tissue-specific manner. The mechanisms by which this is achieved are the subject of our investigations. One classical approach to this problem is to study in depth the elements responsible for a specific gene's expression in a limited subset of tissues. We have taken this approach, using the mouse α-fetoprotein (AFP) gene as the model.

Over several years we have identified the cis-acting regulatory elements that activate transcription of this gene in only three cell types in the mouse—the fetal liver, fetal gut, and visceral endoderm of the yolk sac. These include three 200-base-long segments of DNA that exhibit properties of tissue-specific cellular enhancers and lie substantially upstream of the gene. The fourth positive element lies within the first 125 bp upstream of the gene and participates in the tissue-specific activity of the gene's promoter by interacting with a liver-specific transcription factor, HNF-1.

The AFP gene undergoes a dramatic decline in transcription after birth; this decline is mediated by a fifth cis-acting element that lies between the most proximal enhancer and the promoter of the gene. We have used germline transformation to show that this negatively acting element is activated at birth and interferes with the continued presence of positive transcription factors, not only in the context of the AFP gene, but when it is placed in the same position within the albumin gene, which is not normally repressed after birth. The demonstration that the negative repressor must lie between these strong positive elements to function explains why only the AFP, and not the nearby albumin gene, is negatively regulated in neonatal liver.

The decline in hepatic AFP mRNA after birth is influenced by *raf*, an unlinked gene. Mice that carry a mutant allele of *raf* exhibit 10- to 20-fold higher levels of AFP mRNA in the adult. We used transgenic mice to establish that the *raf* gene product does not act through the negative transcriptional apparatus but rather affects the concentration of AFP mRNA post-transcriptionally, by changing the stability and/or the processing of its transcript. Thus the 10,000-fold decline in AFP RNA after birth is achieved by changes in both transcription and stability of the AFP transcripts.

Our long-standing interest in liver transcriptional repression led us to examine another fetal-specific gene, the *H19* gene. Initially we were surprised to see that this gene contains no long open reading frame, normally found in genes that encode protein products. We have subsequently shown that the RNA is not associated with polyribosomes in the cell but is sequestered in a particle in the cytoplasm. This leaves us with the challenge of understanding the role of a gene transcribed and processed as a classical RNA polymerase II gene, whose product is developmentally regulated in both endoderm and mesoderm, but which does not encode a protein.

One means of understanding the function of any gene is to generate both dominant and recessive mutations in it. To obtain dominant mutants in the mouse, we have introduced excess copies of the *H19* gene into the mouse germline, using microinjection into zygotes. No transgenic mice that express the gene have been recovered at birth. The lethality occurs abruptly, after 14 days into the normal 20-day gestation period of the mouse.

Insight into the basis of the lethality caused by the extra copies of the *H19* gene came from its location on the distal tip of chromosome 7. This region has been shown to undergo genomic imprinting—when both copies of this region are inherited from the mother, the animals die late in gestation. By taking advantage of differences in the *H19* gene between different species of mouse, we showed that only the mother's copy of the *H19* gene is active in a normal mouse. By introducing extra copies of the gene into the germline of mice, we had altered the carefully

controlled gene dosage, and the result was lethality.

Our laboratory is also developing techniques to study the large number of mutations affecting development that have been described in the mouse. Few of these have been approached at the molecular level, primarily because there were no experimental tools available to gain access to the mutated genes. The advent of techniques that allow the investigator to map and clone megabase amounts of DNA at a time—specifically pulsed-field gel electrophoresis and yeast artificial chromosomes—has eliminated this constraint. We have begun a long-term project to develop techniques that will allow us to clone such genes and to establish their identities by complementation analysis via introduction of yeast artificial chromosomes into embryonic stem cells.

A complete library of mouse DNA in yeast, currently under construction, is being used to isolate the mouse *Fused* gene, which affects neural ectoderm development in the early mouse embryo. A 650-kb segment of DNA has been cloned near the *Fused* locus, and a large interspecies backcross is being used to localize its relationship to the *Fused* gene itself. To prove that the *Fused* gene has been identified, large segments of DNA containing it will need to be introduced back into the mouse germline. Strategies for accomplishing this through embryonic stem cells are in progress.

Mechanisms of Gene Regulation in Animal Cells

Robert Tjian, Ph.D.—Investigator

Dr. Tjian is also Professor of Biochemistry 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 Pfizer Award for Enzymology and the Monsanto Molecular Biology Award of the National Academy of Sciences. 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 biochemical process called transcription, which subsequently leads to the production of specific proteins in animal cells. We have devised various means of isolating the individual components of the cell responsible for transcription and have reconstructed this complex reaction in the test tube. In this way we can study how specific genes are turned on and off during cell growth and development. The mechanisms that govern the switching on and off of genes are of fundamental importance in understanding the normal metabolic processes that maintain and perpetuate living cells, as well as in deciphering the basis of disease and other cellular or genetic disorders.

Biochemical Analysis of Cancer Genes

A living cell contains hundreds of thousands of protein molecules, each carrying out its proper function. If either the production or the action of such proteins is altered, severe malfunction can result, such as uncontrolled growth leading to cancer. Among a normal cell's constituents, certain key molecules have the potential to be oncogenic, or tumor causing. Such molecules are called oncoproteins. Our group previously isolated a family of rare proteins from human cells that have been shown to be oncoproteins, encoded by the genes *jun* and *fos*. These regulatory proteins are normally involved in controlling the action of many other genes in the cell, 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 the *jun* 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. The isolation and characterization of this specific negative regulator should help unravel the molecular signaling pathways responsible for transducing information from the outside of the cell into the nucleus, where gene expression is controlled. Such inhibitors of *jun* activity may 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 these rare and fragile proteins. Through the use of specific DNA–affinity chromatography procedures pioneered in our laboratory, we can now isolate transcription proteins that in turn enable us to clone the genes that encode these important regulatory proteins. The ability to isolate this biologically important class of genes provides a powerful approach for studying their structure and function.

In the past several years our 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. These studies are beginning to reveal new concepts regarding the surprisingly modular construction of these 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. Our findings provide the theoretical basis for analysis of other as yet undiscovered transcription factors and will greatly aid our ability to decipher their mechanisms of action.

How Promoter-Specific Regulators Trigger Transcription

One fundamental mystery 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, we recently fractionated and isolated the multiple components necessary to reconstitute transcription. In the process of dissect-

ing the general transcriptional apparatus, we discovered two 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 we anticipate that they will be members of a diverse and essential class of regulatory proteins.

Transcription of Developmentally Regulated Genes

One of our long-term interests is to understand mechanisms underlying the regulation and expression during 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 melanogaster* genes, including alcohol dehydrogenase (Adh), *Ultrabithorax* (*Ubx*), *Antennapedia* (*Antp*), dopa decarboxylase (Ddc), and *hunchback*. 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.

A second interesting finding arising from our studies of *Drosophila* is 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. In addition, the use of fruit flies allows us to probe the developmental and tissue-specific function of this neurogenic regulator in a rapid and highly informative manner not readily accessible in mammalian cells. These studies are expected to yield new insights concerning the tissue-specific distribution and temporal timing of expression during development.

Development and Function of T Cells

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.

THE most critical step in the vertebrate immune response is the recognition of antigens by lymphocytes. This task is accomplished by two sets of glycoproteins: immunoglobulins (Igs) and T cell antigen receptors (TCRs). The most extraordinary feature of these proteins is their structural variability, much of which originates from the ability of the encoding gene segments to undergo somatic rearrangement.

All TCRs were initially thought to be composed of a heterodimeric protein composed of α- and β-subunits. However, the search for the genes encoding these polypeptides led to the identification of a third rearranging gene, γ, that was later shown to encode one of the two subunits of another heterodimeric, TCR $\gamma\delta$. The characteristics of this new type of TCR and the development and function of the T lymphocytes that express it, $\gamma\delta$ T cells, have been the major focus of our research during the past several years.

The two types of T cell—those expressing TCR $\gamma\delta$ and TCR $\alpha\beta$—most certainly arise through two separate cell lineages that originate from a common stem cell. The molecular mechanism wherewith the two cell lineages segregate from the stem cell is of considerable interest, not only for immunologists but also for cellular biologists in general. Our previous studies suggested that there is a DNA element, called a silencer, associated with a γ gene through which its expression is repressed in $\alpha\beta$ T lineage cells. Our data also suggested that the fate of a given precursor T cell is determined by whether or not it contains factors that can specifically interact with the silencer element and thereby repress the associated γ gene.

We have now identified and characterized the silencer element—in fact, two elements, one located within and the other 3′ to a γ gene. Each of these elements alone can repress γ gene transcription in an $\alpha\beta$ lineage cell but not in a $\gamma\delta$ lineage cell, and the two elements, where present in the same $\alpha\beta$ lineage cell, work synergistically. We have also identified a third DNA element between the two silencer elements that enhances γ gene transcription, called an enhancer. We suspect that each of these transcriptional elements (silencers and enhancer) interacts with an array of factors, some of which are differentially present in $\alpha\beta$ and $\gamma\delta$ T cell lineages, and that the differential synthesis or activation of these factors is the key for the segregation of the two T cell lineages.

One of the most intriguing features of $\gamma\delta$ T cells is the compartmentalization of their subsets with different TCR repertoires to various epithelia. Using transgenic mice that were produced by the microinjection of a pair of rearranged TCR γ and δ genes into the pronuclei of fertilized eggs, we have shown that a given $\gamma\delta$ T cell "homes" to a particular epithelium, being guided by a mechanism independent of the specificity of its $\gamma\delta$ TCR. Furthermore, using mice whose thymuses had been surgically removed, we confirmed our earlier conclusion obtained with thymus-less "nude" mice: Development of $\gamma\delta$ T cells that home to spleen and lymph nodes is thymus dependent, while that of $\gamma\delta$ T cells associated with gut epithelium is not.

Another intriguing feature of the compartmentalized $\gamma\delta$ T cells is that virtually no TCR diversity is present among the members of some of these subsets, namely the subset associated with epidermis (called skin-associated intraepithelial lymphocytes, or s-IEL) or the subset associated with the epithelial layer of vagina, uterus, and tongue (called vut-IEL). The structural homogeneity of the $\gamma\delta$ TCR in these tissues is all the more striking because their genes, like Ig and TCR α and β genes, have the capacity for diversity and because this capacity is indeed utilized in some other parts of the body, such as spleen and lymph nodes.

By culturing fetal thymus outside the body in the presence of a monoclonal antibody that will specifically bind to $\gamma\delta$ TCR, we have now demonstrated that the remarkable homogeneity of $\gamma\delta$ TCR of s-IEL and vut-IEL is a result of the strong selection of T cell clones expressing the particular $\gamma\delta$ TCRs through their interaction with un-

known determinants expressed in fetal thymuses (called positive selection). While neither the role of s-IEL or vut-IEL or the nature of the selecting determinants (called ligands) is known, such strong cellular selection of T cells with a particular TCR structure (and hence specificity) suggests its pivotal role in the functioning of these T cells in the respective peripheral site.

We previously showed that at least part of the determinant recognized by a spleen-type $\gamma\delta$ TCR (KN6 $\gamma\delta$ TCR) is encoded by a novel gene ($T22^b$) that maps in the major histocompatibility complex (MHC) and has structural similarity with the MHC class I genes (the products of which are known to present antigen-derived peptides to $\alpha\beta$ TCR). This and related findings by others suggested that the products of a few dozen class I genes clustered in the so-called *Tla* subregion of MHC, whose function has been unknown, may generally serve as peptide-presenting molecules for $\gamma\delta$ TCR. Our recent finding that the product of one of these class I genes, $T3^b$, is expressed, apparently exclusively, on the surface of gut epithelia cells supports this hypothesis. (Note that a particular $\gamma\delta$ T cell subset referred to as i-IEL is localized in gut epithelium.)

While we have accumulated a substantial amount of information with respect to development, diversity, and tissue localization of $\gamma\delta$ T cells, their role in the immune system remains elusive. To help shed light on this most fundamental issue, we have been producing mutant mice deficient in $\alpha\beta$ or $\gamma\delta$ T cells. We accomplish this by manipulating a TCR gene in a test tube, injecting it into the nuclei of a cultured embryonic stem (ES) cell, returning the DNA-injected stem cell into an early embryo (a blastocyst), and finally injecting the embryo into a pseudopregnant mother. The chimeric newborn are tested by breeding them further for germline transmission of the engineered TCR gene, and the germline-transmitted mice are bred by sister-brother mating to generate mice with homozygous mutation. (The entire procedure is called the ES cell gene-targeting technique.)

To date we have succeeded in obtaining mice in which a mutated TCR β, α, or δ gene is germline transmitted. We intend to analyze the immune system of these mutant mice before and after various immunizations, after grafting with normal and tumor cells, and after infecting them with a variety of pathogens. We expect these studies to help not only in identifying the roles of $\gamma\delta$ T cells but also in redefining the role of some $\alpha\beta$ T cells.

Finally, we have also been applying the ES cell gene-targeting technique to a calcium/calmodulin-dependent protein kinase. This kinase is highly enriched in the postsynaptic density of hippocampus and is suspected to be involved in establishing and maintaining long-term memory in mammals. Our goal is to determine whether and how this kinase is involved in memory.

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 School of Medicine, University of California, 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 honors include the Searle Scholars Award and the Passano Foundation Young Scientist Award.

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, 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 optical monitoring is inherently good for following multiple events in parallel, it would be a major help in analyzing the workings of the brain, the most awesome and complex molecular assembly known.

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 sensing mechanisms for 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, we have attached fluorescent labels to 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 neighboring cells 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 rises and falls. While it is in the nucleus, it is ideally placed to modify gene expression.

A complementary area of interest is the use of light both to visualize intracellular biochemistry and to perturb it in a controlled manner to see how the cell or tissue responds. Light is a stimulus that is wonderfully controllable in space and time. Of course most cells, other than specialized tissues like the retina, are not particularly sensitive to light, so the trick is to design and synthesize light-sensitive organic molecules that, introduced into cells, release important messenger substances upon illumination. Through the work of many researchers, ourselves included, such photochemical releaser molecules are available, not only for Ca^{2+} but also other important biochemical signals such as cAMP, inositol phosphates, and diacylglycerol. We have also recently developed the first molecule that can do the opposite function—that is, gobble up the messenger upon illumination. All these photochemical perturbations are valuable to show whether the

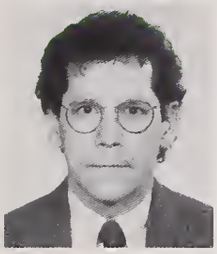
biochemical signal is merely coincidental or is necessary or sufficient for various cell responses.

One problem is that the photochemistry currently used is irreversible, so that once each molecule has been photolyzed and has released or absorbed its cargo, it cannot be recycled back to its initial state. Therefore the experiment is soon over, because the molecules have been used up. The next challenge is to find reversible photochemistry in which light of one wavelength causes messenger release, and light of a different wavelength triggers reuptake. By alternating these two colors of illumination, one would then be able to generate artificial oscillations in messenger level whose effects could be compared with those of the many known natural oscillations. Furthermore, steady illumination with adjacent spots of the two wavelengths should create a standing spatial gradient of messenger level, which would help test whether analogous natural gradients are important in controlling the direction of cell movement and outgrowth, as often hypothesized.

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 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. Possible approaches include microinjection of peptides whose fluorescence is altered by phosphorylation, or development of membrane-permeant fluorogenic substrates for reporter enzymes whose nucleotide sequences can be fused to genes or promoter sequences of interest.

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.

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, and Molecular Biology and Genetics 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 the 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 causative mutations.

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 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. As such, GA serves as a model for more common inherited retinal degenerations whose biochemical basis remains enigmatic. We have been involved in a variety of GA studies, including the development and testing of potential therapies, investigation of possible explanations for the sensitivity of the retina to the systemic biochemical abnormalities, elucidation of regulatory mechanisms controlling the expression of the OAT gene, and determination of the molecular defects causing GA. The last depends on the results of earlier studies in our laboratory, including isolation and sequencing of a full-length cDNA clone of human liver OAT and determination of the structure and organization of the human OAT gene.

The OAT-catalyzed reaction is an essential step in the pathway that interconnects the urea cycle with the tricarboxylic acid cycle. Thus it is not surprising that the regulation of OAT is complex and, in liver, coordinated with other urea cycle–related enzymes. We recognized a sequence motif in the 5′-flanking region of OAT that is similar to sequences in the 5′-flanking region of several other urea cycle–related genes, including ornithine transcarbamylase, argininosuccinate synthetase, arginase, and glutamine synthetase. We designated this candidate cis-acting regulatory element a urea cycle element (UCE). Expression studies with hybrid constructs of various portions of the 5′-flanking region of the OAT gene linked to a reporter gene indicate that the UCE motifs have functional roles. DNase I footprint assays and gel retardation experiments with a series of synthetic oligonucleotides corresponding to the normal and mutant variations of the UCE sequence indicate that at least three proteins, each with distinct but overlapping recognition sites, bind to the UCE.

Using a concatamer of the UCE sequence as a radiolabeled probe, we have cloned cDNAs for two specific DNA-binding proteins that are candidates for trans-acting regulators of the homeostatic mechanisms that adjust ureagenic capacity to nitrogen intake. Transient transfection experiments in which we co-transfect one of these cDNAs with a second construct containing two or four UCE in the promoter of a reporter gene indicate that the protein product of our cDNA clone does activate the expression of genes with UCE in their promoters.

To better understand the coordinated regula-

tion of the urea cycle–related enzymes, we recently have embarked on a project to clone the genes for additional enzymes that are metabolically related to OAT. For several of these, purified protein or sequence data are not available; therefore we are cloning these genes by functional complementation of mutant variants of *Saccharomyces cerevisiae*. To date we have cloned a full-length cDNA from human pyrroline-5-carboxylate reductase by this method. Preliminary studies support the feasibility of using the same strategy to clone pyrroline-5-carboxylate dehydrogenase. These two enzymes catalyze reactions involving the product of the OAT reaction, and the latter is the site of the primary defect in the human disorder, type II hyperprolinemia. We anticipate that the genes for these enzymes will be subject to transcriptional regulation that interacts with that of OAT.

We are continuing an extensive molecular analysis of the OAT mutations that cause GA in the probands of 79 families from around the world. The functional consequences of the detected mutations are tested by expressing the mutant allele in a Chinese hamster cell line that lacks endogenous OAT protein or mRNA. Of the 27 alleles we have detected, 15 have been confirmed as deleterious, 10 are presumed to be deleterious, 1 is neutral, and 1 is synonymous. In addition to providing a means for accurate molecular diagnosis of GA, we anticipate that our examination of these defects, particularly the 18 missense mutations, will enhance our understanding of OAT structure and function. We are exploring strategies to express OAT in quantities sufficient for crystallographic analysis of OAT structure.

Our studies of the mutations causing GA have highlighted the remarkable allelic heterogeneity at the OAT locus. With one exception (L402P, the most prevalent allele in Finnish GA patients), each mutant allele accounts for less than 10 percent of the total. Survey of the genotype of probands from our 79 GA pedigrees indicates that the 35 mutant alleles so far discovered by ourselves and others account for only 70 percent of the total mutant OAT genes in this population.

A final OAT-related project depends on the utilization of a cluster of OAT-related sequences (all located on the proximal portion of the short arm of the X chromosome) to develop a large-scale genetic map of this region of the human genome. The genes for several diseases involving the retina map to this region. At this point we have cloned and organized two regions, each about a megabase, in Xp11.2-Xp11.3, which are separated by a 1–3 megabase interval. We are beginning to identify the active genes in these areas and will examine their possible role in inherited retinal degenerations.

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 tissue from these patients exhibit deficiency of virtually all peroxisomal enzymes. In collaboration with Hugo Moser, we are cloning the genes for integral membrane proteins unique to peroxisomes. The first of these, the 70-kDa peroxisome membrane protein (PMP), is a member of the multiple-drug resistance (MDR) superfamily of proteins, which also includes the CFTR protein involved in cystic fibrosis. We have cloned and sequenced a full-length cDNA for the human PMP and are in the process of determining if mutations in this gene are responsible for any of the inborn errors of peroxisomal biogenesis.

The Molecular Physiology of Calcium

D. Martin Watterson, Ph.D.—Investigator

Dr. Watterson is also Professor of Pharmacology at Vanderbilt University School of Medicine. He received his Ph.D. degree in biochemistry from Emory University. He was then an NIH postdoctoral fellow at Duke University Medical Center. Before joining the faculty at Vanderbilt, Dr. Watterson was Assistant Professor (Andrew Mellon Fellow) and Associate Professor of Cell Biology at the Rockefeller University.

THE maintenance of a viable basal state in eukaryotic cells (homeostasis), and the responses of cells to environmental stimuli, involve a fine-tuned regulatory network that is modulated by transient changes in the intracellular levels of ionized calcium. The molecular mechanisms by which quantitative changes in calcium concentration are transduced into qualitative changes in cell behavior (altered cell motility, proliferation, receptor capping, cell morphology, etc.) involve the reversible interaction of calcium with a class of protein receptors that bind it selectively in the presence of higher concentrations of other ions, such as magnesium. A prototypical member of this class of calcium-binding proteins is calmodulin.

Calmodulin is ubiquitous among the plant and animal kingdoms and has multiple biological roles. It fulfills these roles through its presence as an integral subunit of a diverse array of enzymes, cytoskeletal structures, and membrane-associated transport systems. Insight into the fundamentals of this regulatory network would provide a framework from which to interpret biological phenomena and guide clinical treatments. It is the goal of our research to elucidate the calmodulin-mediated signal transduction system and thereby gain a more complete understanding of how calcium regulates cell processes and how this homeostatic network might be altered in certain disease states or susceptibilities.

Calmodulin-regulated Protein Kinases

One class of enzymes that have calmodulin as an integral calcium-binding regulatory subunit are the protein kinases. Protein kinases catalyze the transfer of phosphate from an ATP molecule to a specific site on certain proteins (*phosphorylation*), resulting in a rapid alteration of the phosphorylated protein's biological functioning.

In the basal or "resting" state of the cell, the active site of the protein kinase is inhibited by another region of the kinase molecule (*autoinhibition*). When calcium binds to calmodulin, which is loosely bound to the kinase, the calmodulin's three-dimensional structure is subtly changed. This in turn results in a change in the three-dimensional arrangement of the protein kinase molecule that relieves the inhibition of the kinase active site. The kinase can now phosphorylate a specific set of proteins in the cell and thereby set off a cascade of events.

The basal state is reestablished by at least two closely coupled reactions: the removal of phosphate from the phosphorylated protein by an enzyme called a phosphatase, and the dissociation of calcium from calmodulin with a resultant restoration of the protein kinase to its autoinhibited state. During this past year, we elucidated the primary structure of one of these calmodulin-regulated protein kinases—called nonmuscle myosin light-chain kinase (nmMLCK)—that is found in all nonmuscle tissues examined.

We have also determined, through use of molecular genetic and protein chemistry technologies, some fundamental features of both the protein kinase and calmodulin that are required for these two proteins to recognize each other in a milieu of macromolecules, with the resultant formation of a fully functional calcium signal-transduction complex.

By serendipity, we also discovered a novel genetic relationship among this protein kinase and some other proteins produced from the same gene. As a result, we have extended our investigations to include an analysis of how the various proteins with different biological functions are made from this single gene and a more detailed localization of the gene in the human genome.

Calmodulin-regulated Ion Channels

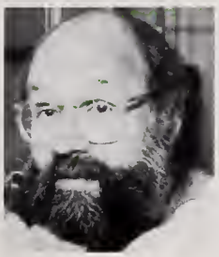
Ion channels are cell membrane proteins that regulate ion transport across the membrane. They do not appear to be enzymes like the protein kinases, but represent another class of cellular proteins whose functions can be regulated by calmodulin. Our previous studies of mutant organisms with ion channel defects demonstrated that it is possible to have inherited mutations of calmodulin that can selectively alter one calmodulin-regulated pathway. (Previously it was thought that any inherited mutation of calmodu-

lin would be lethal.) During this past year, we continued our studies of the structural features of calmodulin required for proper targeting to cell membranes and for regulation of ion channel–mediated cell behavior. These studies showed that a single amino acid, unchanged in calmodulins from plant and animal species, is required for full ion channel regulatory activity.

A spin-off from this study was the production of new proteins, by using designer (modified) genes, that are useful diagnostic reagents for future explorations of the molecular mechanisms involved in ion channel regulation. Knowledge about how calmodulin affects the activity of ion channels will provide insight into the features of another class of proteins and regulatory pathways modulated by calcium through calmodulin, thus complementing the knowledge gained from studies of the protein kinase family of calmodulin-regulated proteins.

The fact that protein kinases are modulated by ions whose flux is regulated by ion channels, and that ion channels can be regulated by protein kinase–catalyzed phosphorylations, indicates the potential of studies focused on selected calmodulin-regulated proteins for developing a cohesive model of the feedback and cross talk that occurs *in vivo*.

MyoD: *A Master Regulatory Gene for Myogenesis*



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 in the Department of Zoology at the University of Washington. 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. Prior to joining the staff at the Hutchinson Center, Dr. Weintraub was a member of the Department of Biochemical Sciences at Princeton. He is a member of the National Academy of Sciences and the American Academy of Arts and Science.

THE nucleus of a blood or skin cell has the capacity, when put back into an enucleated egg, to give rise to an entire animal, containing perhaps a thousand cell types. This demonstrates that a given cell type contains all the genetic information needed to generate all other types. How, then, are the different cells in the body generated? The answer is that different sets of genes are used to define different cell types—e.g., muscle-specific genes are expressed only in muscle cells, and nerve-specific genes, only in nerve cells. This solution has been confirmed in hundreds of experiments by many investigators over the past 10–20 years.

But how are these different batteries of cell type–specific genes activated in the correct places and times in the embryo? The goal of our laboratory is to answer this question, and we have chosen to ask specifically how muscle cells are generated during development.

MyoD is a Master Regulatory Gene for Myogenesis

Our initial experiments identified a gene, *MyoD,* by the criteria that it was expressed in myoblasts but not in fibroblasts, nor in mutant muscle cell lines that had lost the capacity to differentiate into muscle. Detailed analysis showed that MyoD is expressed only in skeletal muscle cells and their committed myoblast precursors. When *MyoD,* however, is introduced into fibroblasts using a viral long terminal repeat (LTR) to promote constitutive expression, these cells activate the myogenic program and become muscle. A large number of other differentiated cell types —fat, pigment, brain, cartilage, etc.—can be converted to muscle by LTR-induced expression of *MyoD.* In the process, the normal program of these cells is usually turned off.

We consider *MyoD* a "master regulatory gene" for myogenesis and view it as a "nodal point" in the flow of myogenic information. Many temporal and spatial cues from the early embryo impinge on *MyoD* to turn it on or off, but once activated the gene can keep itself on (since it promotes its own transcription) and then turn on (directly or indirectly) the entire battery, perhaps hundreds, of muscle-specific genes.

MyoD Protein Is a Transcription Factor That Activates Muscle-Specific Genes

MyoD is a nuclear protein, 318 amino acids in length, that binds to muscle-specific enhancers. Many, if not most, muscle enhancers contain multiple MyoD-binding sites. MyoD binds to a consensus binding sequence of which the apparent simplicity belies a rather sophisticated interplay among protein, DNA, and subsequent cell type–specific transcriptional activation.

Only 68 amino acid residues of MyoD are required for stable muscle cell conversion. This region contains the DNA-binding domain, which is a putative basic-helix-loop-helix (bHLH) structure. There are now well over 30 known bHLH proteins, sometimes referred to as the *Myc* homology family. Extensive mutagenesis has revealed that the HLH region is required for dimerization, and the adjacent basic region, for DNA binding.

The Activity of the MyoD Protein Is Regulated

Proliferating myoblasts in tissue culture express MyoD RNA and protein, yet do not differentiate. As assayed by *in vivo* footprinting, the MyoD-binding site in the enhancer of the *MCK* gene is occupied in differentiated myotubes but not in myoblasts. Thus there are physiological controls that regulate MyoD activity. The decision either to proliferate or to differentiate into myotubes is governed by the presence or absence of serum, but the mechanism of action is unknown. One attractive model is that a second type of HLH protein is controlling MyoD activity. We have been investigating this possibility.

Information Processing by HLH Proteins

Several types of biologically important decisions use information provided by HLH proteins. Two of these, *MyoD* and the *achaete-scute* product (an HLH protein important for neurogenesis),

seem to provide information about cell type. *Id* (inhibitor of differentiation) offers a pathway for negative control. And *E2A* may perform an effector or integration function yet to be defined fully. These types of inputs then seem to be integrated by HLH interactions between pairs of individual monomers, and the result seems to be a specific binding capacity. For protein complexes containing Id, the choice is probably on-or-off; for complexes of E2A or MyoD homodimers or heterodimers, the result can be altered DNA-binding specificity. Although it is clear that interactions between HLH proteins are very precise and specific, the rules are not yet known.

MyoD Is Controlled by and Can Control the State of Cell Growth

A variety of agents specifically inhibit myogenic differentiation. The list includes genes such as *ras*, *fos*, *jun*, *fps*, *erbA*, *myc*, and *E1A*; chemical agents such as butyrate and phorbol esters; and growth factors such as FGF and FGFβ. Most of these reagents can inactivate the expressed MyoD protein, and several, such as *ras* and *fos*, also inhibit *MyoD* transcription. The specific pathway by which each of these oncogenes, anti-oncogenes, and growth factors inhibits myogenesis provides a clue to how MyoD might integrate information coming from many aspects of cellular function.

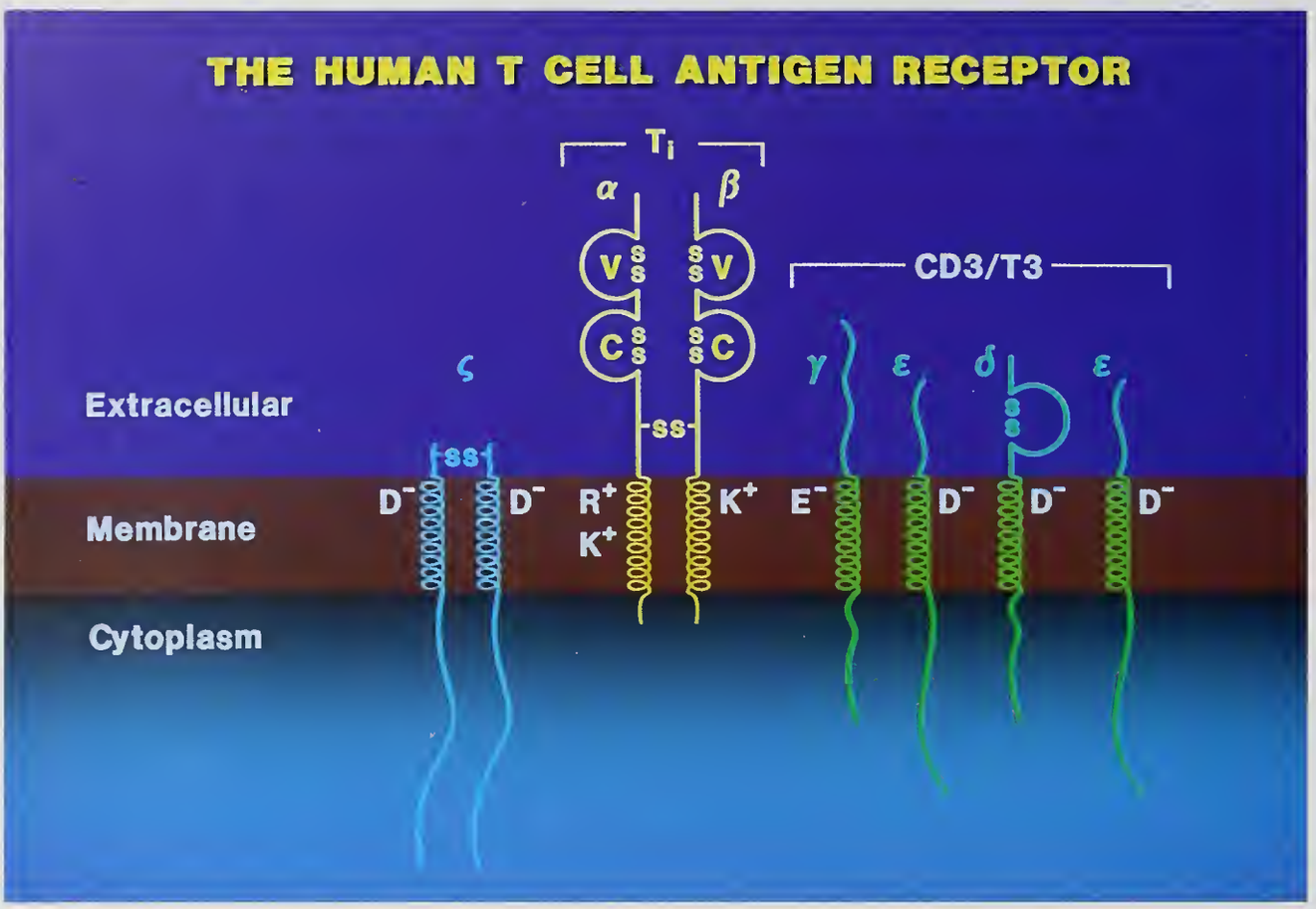
Factors that influence growth regulate MyoD, but MyoD itself can also inhibit cell growth. Experiments with MyoD mutants show that the HLH domain is required for growth arrest.

Activation of MyoD During Development

Recently, the *MyoD* gene has been cloned from *Caenorhabditis elegans*, and an antibody to the protein, produced. The antibody first stains cells at the 100-cell stage. These are lineage founder cells fated to give rise to body wall muscle descendants some three to four generations later. Less than 3 kb of the upstream control region of MyoD is needed for body wall muscle-specific expression, as revealed by fusing these sequences to a β-galactosidase expression reporter. Results suggest that the capacity for transcription of *MyoD* in a myogenic lineage may precede the actual cell type–specific expression of the protein, and this capacity segregates into specific blastomeres as early as the 28-cell stage. Presumably, subsequent cell-cell interactions determine which descendants will give rise to nerve and skin and turn off MyoD transcription and which will maintain MyoD expression and give rise to skeletal muscle cells.

Conclusion

The flow of cell type–specific information for myogenesis from the egg to the final muscle cell goes through a nodal point, defined by functions supplied by the *MyoD* family of myogenic regulatory proteins. It is possible that the *achaete-scute* complex of genes also acts as a nodal point, but for neurogenesis. An alternative organization could have also been imagined: combinations of less tissue-restricted genes could have encoded the *MyoD* or *achaete-scute* functions. The nodal point need not have been an optimal solution, but simply one that worked at some crucial time during evolution. Whether other cell lineages employ this strategy, and whether there is a profound insight provided by its use (such as the potential to make cell types mutually exclusive), await further analysis.



Model of the structure of the human T cell antigen receptor. Basic and acidic amino acid residues present in the transmembrane domains of each of the chains of the receptor are indicated by the one-letter amino acid code (D^-, aspartic acid; E^-, glutamic acid; R^+, arginine; K^+, lysine). V and C of the Ti chains correspond to variable and constant domains within these immunoglobulin-like chains involved in antigen recognition. The ζ-chain dimer and the γ-, δ-, and ε-chains of the CD3/T3 complex are believed to function in signal transduction.

Research of Arthur Weiss.

Structural and Functional Studies of the T Cell Antigen Receptor

Arthur Weiss, M.D., Ph.D.—Associate Investigator

Dr. Weiss is also 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 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 a 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. The 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 the six invariant chains of the CD3 complex. Ti may be viewed as the ligand-binding subunit of the TCR, since it contains all the information needed to recognize antigen and MHC specificities. CD3 has 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 CD3 and Ti is of interest.

Previous studies from our laboratory have demonstrated that coexpression of CD3 and Ti on the plasma membrane is obligatory. In recent studies, we have found that the structural and functional basis for the interaction between Ti and CD3 is contained within the transmembrane domains of these proteins. Further mutational studies are in progress to understand more precisely how Ti and CD3 interact functionally within these domains.

Since the transmembrane regions are responsible for the Ti-CD3 association, we have taken advantage of this information to separate regions or domains of the CD3 chains from Ti. This has been accomplished by constructing chimeric molecules between other cell surface molecules linked to the cytoplasmic domain of the CD3 chains. Chimeric molecules including the CD3 ζ-chain acquire the signal-transducing capacity of the entire TCR. These results demonstrate that the cytoplasmic domain of CD3 ζ links the TCR to intracellular signaling machinery. The ability to create such functional chimeric receptors may permit the creation of new antiviral or antitumor TCRs, which may be valuable in gene therapy.

The interaction of the TCR with its ligand, antigen/MHC, initiates cellular activation by inducing a transmembrane signal. Such signal transduction is manifested as the formation of intracellular biochemical mediators called second messengers, which can initiate or influence cellular response pathways. The TCR activates two signal transduction pathways. One signal transduction mechanism, the inositol phospholipid pathway, involves receptor-stimulated hydrolysis of a rare membrane lipid, PIP_2 (phosphatidylinositol 4,5-bisphosphate). This yields two potent intracellular second messengers (inositol 1,4,5-trisphosphate and diacylglycerol), which regulate the mobilization of the enzyme protein kinase C. These latter two events are physiologically important to subsequent cellular responses.

The TCR activates a second signal transduction pathway that involves an enzyme with tyrosine kinase activity. The importance of the kinase pathway and its relationship to the inositol phospholipid pathway are being clarified by recent studies. We have found that stimulation of the TCR induces the tyrosine phosphorylation of an isozyme of phospholipase C, an enzyme that catalyzes the hydrolysis of PIP_2. These studies, together with additional studies using tyrosine kinase inhibitors, suggest that activation of a tyrosine kinase that is regulated by the TCR is required for activation of the phosphatidyl inositol pathway.

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. 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 Ti 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. The defect in another of these mutants can be attributed to the absence of CD45, a cell surface protein with tyrosine phosphatase activity. The absence of CD45 prevents the TCR from activating the tyrosine kinase or phosphatidyl inositol pathway. This suggests a complex autoregulatory system, which we are intensively studying. In the remaining two mutants, biochemical studies have complemented our genetic approach and have suggested potential defects. We previously identified two molecules that associate with the TCR in a ligand-induced manner. In these two mutants, these molecules fail to associate with the TCR. Further studies of the basis for the defective signal transduction observed in these mutants are in progress. 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.

An alternative approach toward understanding which signal transduction pathways can lead to cellular responses is to express other heterologous receptors with well-defined signal transduction functions in T cells and examine their abilities to activate typical T cell responses. For example, the human muscarinic acetylcholine receptor subtype 1 (HM1) activates the inositol phospholipid pathway in cardiac, smooth muscle, and neuronal cells. We have expressed HM1 in a T cell line and find that, when stimulated, it can activate the inositol phospholipid pathway but not a tyrosine kinase pathway. Moreover, HM1 stimulation results in lymphokine production and other typical results associated with TCR stimulation. These results suggest that the inositol phospholipid pathway, a signal transduction pathway common to both the TCR and HM1, does regulate T cell responses. How this pathway serves to regulate later cellular responses is being addressed in ongoing studies.

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 cell in the bone marrow that has multipotent capacity.

Three years ago we were able to isolate the hematopoietic stem cell of the mouse. This past year we demonstrated its full developmental potential by transferring a single stem cell from one strain mixed with 100 stem cells from another strain of mouse into lethally irradiated mice of the second strain. Remarkably, in a significant fraction of animals, the stem cells gave rise to blood cells and tissue lymphocytes. In every case where donor-strain cells were found, all blood cell types had come from the single marked cell.

In one-third of these cases, the stem cell proliferated; very large numbers of daughter blood cells were generated through the life of the animal. In these mice, 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 have traced how one of the first daughter cells from the stem cell—the pre–B lymphocyte in bone marrow—becomes leukemic. This work was supported jointly by the National Cancer Institute of the National Institutes of Health and the Howard Hughes Medical Institute. It had its origin several years ago when my laboratory—and independently that of another HHMI investigator, Max Cooper (University of Alabama at Birmingham)—identified a molecule present at low levels on normal pre–B lymphocytes and at much higher levels in leukemias and lymphomas of the pre-B series. We called the molecule the 6C3/BP-1 antigen, here abbreviated Ag (to signify antigen).

Like hematopoietic stem cells, pre–B cells are dependent for their growth on a class of stromal cells found embedded inside bone cavities—the bone marrow. We had cloned the particular bone marrow stromal cells that support normal pre–B lymphocytes and had found that they provide a good culture microenvironment in which to study the process whereby Abelson leukemia virus causes pre–B cells to become leukemias and/or lymphomas. Among the products of these stromal cells is the protein growth factor interleukin-7 (IL-7). Normal pre–B cells don't express IL-7, while the stromal cells that support the growth of these normal and preneoplastic pre–B cells do.

The Ag is only expressed at high levels on these leukemias and lymphomas at a late stage of the neoplastic development in a test tube containing stromal cells, pre–B target cells, and the leukemia virus. We were therefore interested in determining whether the conversion from stromal cell dependence to stromal cell–independent malignant growth resulted from tumor cell expression of IL-7. In fact, every Abelson leukemia virus–induced pre-B leukemia expressed the growth factor. More importantly, by specifically blocking the expression of IL-7, we could prevent the growth of these leukemias in tissue culture.

Thus we have demonstrated what we think is a critical step in the passage from normal cell to leukemic cell—the expression of a growth factor for which the tumor cell had a receptor. This result confirms in part a hypothesis we proposed over 15 years ago: receptor-mediated leukemogenesis.

We next investigated whether the high-level expression of the Ag might have some significance for the growth of tumor cells *in vitro* or *in vivo*. Our collaboration with Max Cooper brought some striking insights to this problem. His group had cloned the cDNA encoding the Ag and had found it to contain sequences in common with aminopeptidase A, a metalloproteinase.

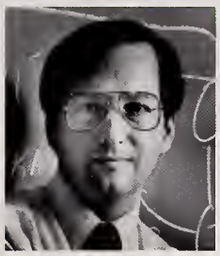
(The protein is an enzyme that can digest certain other proteins.) At the same time we had developed several subclones from a clonal line of Abelson leukemias, some that expressed Ag molecules (Ag^{hi}) and others that did not express them (Ag^{neg}). The Ag^{hi} and Ag^{neg} lines were stable for those properties over tens of cell generations in the test tube.

We wished to test if Ag^{hi} cells might use their cell-surface protease to digest the surrounding tissues, invade adjacent blood vessels, and spread via the bloodstream to distant sites. Our experiments support this hypothesis. Ag^{hi} subclones placed in the thigh muscle grew locally, invaded the bloodstream at our earliest observation interval (6 days), and killed 50 percent of the hosts in 17 days. The same number of Ag^{neg} subclones placed in the thigh muscle grew locally, but no cells could be found in the blood until 17 days, and the 50 percent death point came at 28 days. Thus, although the tumors grew locally at roughly equal rates, only the Ag^{hi} cells invaded the bloodstream early and caused early deaths.

Perhaps most significantly, the bloodstream leukemia derived from Ag^{neg} cells now contained a significant fraction of Ag^{hi} cells—which did not occur when they grew in the test tube. That the early death of animals was due to the high invasive properties of Ag^{hi} cells rather than some intrinsic growth advantage they might possess was made clear when we showed that equal numbers of Ag^{hi} and Ag^{neg} cells injected directly into the bloodstream led to death of leukemic hosts at almost the same time.

We now need to know how reproducible this condition is and whether manipulation of the sequence encoding the Ag will reveal how the enzyme acts in malignant cells to promote their invasiveness and in normal bone marrow differentiating from hematopoietic stem cells to B lymphocytes. And we need to determine whether there is a human homologue active in normal and neoplastic hematopoiesis.

Regulation of Cell Membrane Ion Channels



Michael J. Welsh, M.D.—Investigator

Dr. Welsh is also Professor of Internal Medicine and Physiology and Biophysics at the University of Iowa College of Medicine. 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.

CELL membrane ion channels control the intracellular ionic environment and play an important role in signal transduction. In epithelial cells they are also responsible for secretion and absorption of electrolytes and water.

In the lungs, ion transport, including chloride secretion by the airway epithelium, controls the quantity and composition of the respiratory tract fluid. Thus it helps effect mucociliary clearance, a normal pulmonary defense mechanism.

Net vectorial ion transport requires the asymmetric distribution of ion channels and other transporters in polar epithelial cells. For active chloride secretion, chloride enters the cell at the basolateral membrane (interstitial side of the cell) by a sodium-coupled cotransport process that accumulates intracellular chloride above electrochemical equilibrium. Chloride then leaves the cell passively, exiting across the apical membrane (luminal side of the cell) via a chloride channel.

Intracellular sodium is maintained at a low concentration by the basolateral sodium-potassium ATPase, and potassium that enters on the sodium pump is recycled by basolateral potassium channels. Effective secretion requires coordinated regulation of chloride channels at the apical membrane and potassium channels at the basolateral membrane. The regulation of these channels controls the rate of transepithelial chloride secretion. An understanding of the function and regulation of epithelial ion channels is a major goal of this work.

Understanding the regulation of apical chloride channels is also important in understanding cystic fibrosis (CF), a common lethal genetic disease of Caucasians. In CF the regulation of apical chloride channels is defective. Apical membrane chloride channels are activated and chloride secretion is stimulated by a variety of hormones and neurotransmitters that increase intracellular levels of cAMP. In CF the apical membrane is chloride impermeable, an abnormality that may explain several of the pathologic manifestations. In CF airway epithelial cells, hormonal secretagogues stimulate cAMP accumulation appropriately but fail to activate (or open) chloride channels.

In many cells the biologic effects of cAMP result from activation of cAMP-dependent protein kinase (PKA), resulting in phosphorylation of target proteins. To determine if the enzyme PKA regulates chloride channels, we used excised, inside-out patches of membrane and added the purified catalytic subunit of PKA plus ATP to the internal surface of the channel. We found that chloride channels were activated by PKA plus ATP in cell-free patches of membrane from normal cells but were not activated in CF cells.

Protein kinase C (PKC), another important enzyme, regulates cell function by phosphorylation. In airway epithelia, several secretagogues increase the cellular mass of diacylglycerol, an activator of PKC, suggesting that the latter also regulates the chloride channel. In experiments using excised cell-free patches, we found that PKC could have a dual effect on chloride channels. At a high calcium concentration, it inactivated chloride channels, and at a low calcium concentration, it activated them. In CF cells, PKC-dependent channel inactivation was normal, but activation was defective. Thus it appears that PKC also regulates the chloride channel by phosphorylation and suggests that two different sites on the channel, or an associated membrane protein, may be involved in regulation, with one of these being defective in CF.

Although phosphorylation-dependent activation of chloride channels is defective in CF, an increase in the cell calcium concentration stimulates chloride secretion in airway epithelia. Calcium-stimulated chloride secretion has the potential to bypass the CF secretory defect. We found that calcium and cAMP activate different chloride channels in the apical membrane. This conclusion is based on four observations: 1) calcium-activated chloride channels are present in the apical membrane of airway, but not intestinal, epithelia; 2) cAMP-activated but not calcium-activated chloride channels are defective in CF airway epithelia; 3) calcium- and cAMP-activated channels have different halide permeabilities;

and 4) an increase in both second messengers produces an additive increase in chloride current. These results also explain the previously puzzling observation that calcium-stimulated chloride secretion is defective in CF intestine but intact in CF airway: the calcium-activated chloride channels that could circumvent the chloride secretory defect in CF airway are missing from the apical membrane of intestinal epithelia.

The gene that is defective in CF has recently been cloned by other investigators. This gene —the cystic fibrosis transmembrane conductance regulator (CFTR)—was cloned based on a correct chromosomal location, an appropriate pattern of tissue expression, and a mutation (ΔF508) that was present on 70 percent of CF chromosomes and absent on non-CF chromosomes.

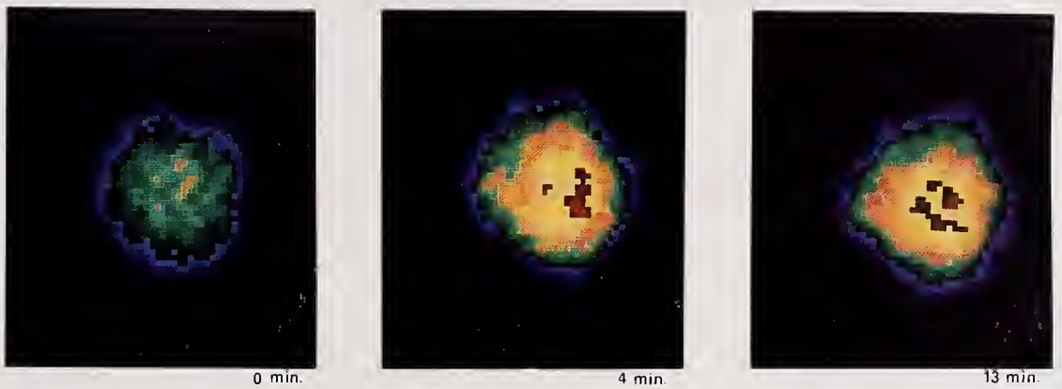
We tested the hypothesis that expression of normal CFTR would complement the CF phenotype, by expressing nonmutated CFTR in cultured CF airway epithelial cells. We assessed chloride channel activation in single cells using a fluorescence microscopic assay and the patch-clamp technique. Expression of CFTR, but not the ΔF508 mutation, corrected the chloride channel defect. These results demonstrate a causal relationship between mutations in the CFTR gene and the CF phenotype. Because expression of mutated CFTR did not complement the CF phenotype, other interpretations are unlikely. Demonstration that the CF phenotype can be corrected in cultured CF cells suggests the feasibility of a therapeutic approach based on correction of the underlying defect.

Although correction of defective chloride transport in CF epithelial cells strengthened the relationship between CF and epithelial chloride channels, it did not demonstrate the function of CFTR. To evaluate its function, we expressed CFTR in several heterologous cells: HeLa cells, Chinese hamster ovary cells, and NIH 3T3 fibroblasts. In each of these cell types, expression of CFTR, but not the ΔF508 mutant, causes an increase in cAMP-stimulated anion permeability and chloride currents. These data show that expression of CFTR produces cAMP-activated chloride currents in cells that do not normally express CFTR and do not normally have cAMP-activated chloride channels. When CFTR is overexpressed in cells that already express high levels of CFTR, the chloride permeability increases even further. These data suggest that CFTR is itself a cAMP-activated chloride channel. Other alternatives have not yet been excluded, but all would require that this diverse group of cells must contain quiescent chloride channels that can become sensitive to cAMP only in the presence of CFTR.

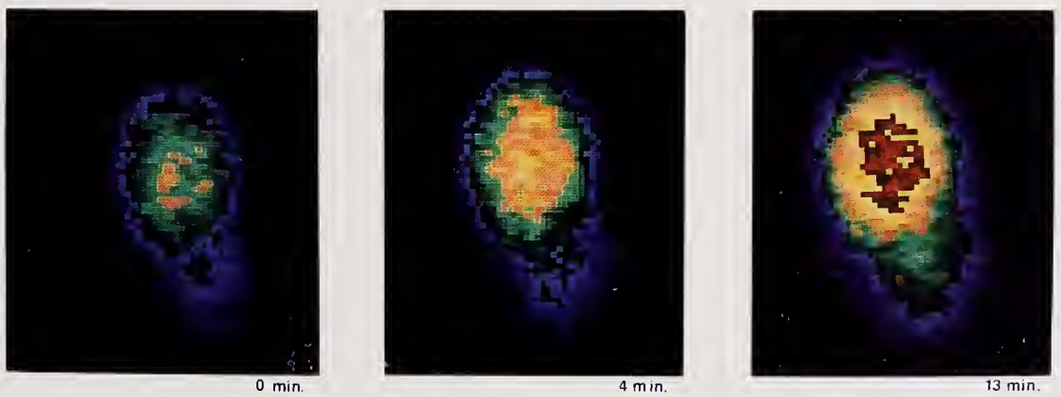
In addition to studies of the chloride channel, the laboratory also focuses on understanding the function and regulation of basolateral membrane potassium channels. The coordinated regulation of basolateral potassium channels and apical chloride channels is critical to effective chloride secretion. Our current work suggests the presence of two potassium channels at the basolateral membrane—one regulated by calcium and one regulated by some other agent. The changes in the activation of the potassium channels are critical to prevent changes in cell volume and to maintain an electrically negative intracellular voltage that drives chloride out of the cell across the apical membrane.

A third aim of the laboratory focuses on the regulation of intracellular calcium concentration, which controls many cellular processes. Stimulation of single cells with agonists often does not produce a steady increase in cell calcium; instead it causes oscillating, or pulsatile, elevations of calcium. Such oscillations have been observed in many cells, and it has been proposed that the frequency at which calcium oscillates determines the biologic response. We have turned to the *Xenopus* oocyte as a model system that allows us to manipulate the factors that control intracellular calcium. Studies of the genesis of oscillations in cell calcium in *Xenopus* oocytes should be relevant to understanding mechanisms that underlie this important regulatory system in many cells.

CFTR

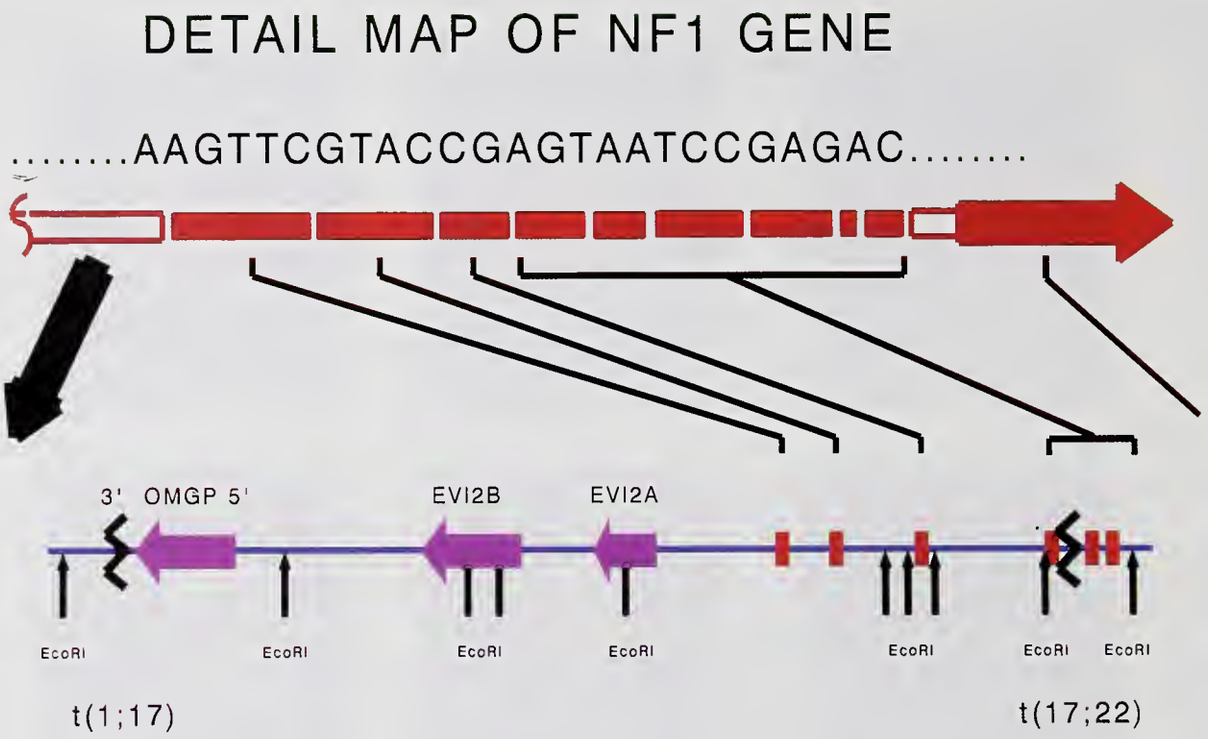


CFTRΔF508



Chloride channel function in lung cells from a cystic fibrosis patient returns to normal following insertion of a normal gene. The gene for cystic fibrosis transmembrane conductance regulator (CFTR) is mutated in patients with CF, and defective function of chloride channels is the hallmark of the disease. The top row shows fluorescence from a cystic fibrosis cell that is expressing normal CFTR; the bottom row shows a cell that is expressing mutated CFTR. Fluorescence indicates chloride channel activity, with yellow and red indicating high fluorescence and blue and green low fluorescence. Fluorescence increases faster in cells expressing the normal CFTR gene (top). The difference between cells expressing the normal gene and the mutated gene is most noticeable at the 4 min point.

Research of Michael J. Welsh.



Structure of part of the NF1 gene (above), in relation to a genomic map of the region (below). The map is oriented with the centromere of chromosome 17 to the left. Sequenced exons, or coding regions, are shown in orange, and the direction of transcription is indicated by the large orange arrow. The points where the translocation occurs in each of two NF1 patients, t(1;17) and t(17;22) (jagged lines), are 50–60 kilobases apart. These points provided major clues to the whereabouts of the NF1 gene.

The three genes known to lie within NF1, and their orientation on the chromosome, are shown in pink. Sites indicated by vertical black arrows mark where the enzyme EcoRI *cleaved the genomic DNA into fragments to form the restriction map.*

Only nine exons of NF1 are shown here. Others exist in either direction (indicated by large arrows), but within this region the sequence of individual DNA base units (represented by a sequence of letters A, C, T, and G) translates into a peptide capable of controlling growth signals within a cell.

Adapted from White, R.L. 1991. Identification of the neurofibromatosis gene. In Origins of Human Cancer: A Comprehensive Review. *Cold Spring Harbor, NY: Cold Spring Harbor.*

Identification and Characterization of the Gene Responsible for Neurofibromatosis Type 1

Raymond L. White, Ph.D.—Investigator

Dr. White is also Professor of Human Genetics 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 has received many honors for his cancer research and recently was co-recipient of both the Charles S. Mott Prize and the National Neurofibromatosis Foundation's Friedrich von Recklinghausen Award.

THE search for the gene that is altered in a dominantly inherited disorder, neurofibromatosis type 1 (NF1), was successful when workers in our laboratory found in patients an abnormal sequence of DNA in the relevant region of chromosome 17. Knowledge of the normal and abnormal functioning of this gene will elucidate the biochemical processes that lead to disease in patients with NF1, and will open new opportunities to develop therapies for this relatively common and sometimes devastating disorder.

Furthermore, the product of this gene falls into the category of proteins that help to control growth and differentiation, and for that reason the discovery has broad implications for those seeking to learn more about the intracellular chemical signals that determine the basic activities of life and that may lead, when defective, to the formation of tumors.

The region of chromosome 17 known to harbor the NF1 gene had been narrowed to a point where only a few genes were likely to exist. However, the first candidate to be examined, *EVI2A*, did not differ in DNA sequence between NF1 and non-NF1 individuals.

Two other genes, *OMGP* and *EVI2B*, were subsequently identified. *OMGP* was especially interesting because it is the coding element for a protein expressed in cells of the central nervous system—cells thought to be counterparts of the peripheral Schwann cells affected in NF1. Moreover, *OMGP* lies very near to the site of a chromosomal break in a patient with NF1, a mutation assumed to be the cause of her disease. Nevertheless, *OMGP* and *EVI2B* were eventually eliminated as candidates on the same grounds as *EVI2A*.

All three genes lie in the same orientation on the chromosome. That is, they are all transcribed, or "read," in the same direction when the genetic code is being translated to form a protein within the cell.

Finally, examination of DNA from many individuals detected three deletion mutations specific to NF1, within the region defined by translocations in two unrelated patients. Two of these deletions of DNA material were large enough to disrupt one or another of the known genes. The third, however, did not affect any of them. This observation suggested that the mutation in that patient was disrupting a fourth gene. Furthermore, the sequence affected by this mutation was very similar to one present in mouse DNA. (Conservation of sequence in evolution between widely disparate species often signals the presence of an important coding element.)

The conserved sequence was used to seek out complementary DNA (cDNA)—that is, recombinant DNA copies of an active gene—from a "library" of clones reflecting gene expression in human fetal brain tissue. When the DNA sequence of the cDNA corresponding to this new gene was compared among normal individuals and several NF1 patients, a number of mutations involving single base pairs of DNA were detected, and at least some of these changes were of a kind expected to be damaging to the protein product. These results, combined with other information, indicated that the NF1 gene had at last been found.

The remarkable finding of three other genes *within* the NF1 gene, all oriented in the opposite direction to it on the chromosome, appears unique so far in human molecular biology. The significance of this phenomenon, and the possible effect(s) of mutations in the NF1 gene upon expression of any of the embedded genes, are subjects for further investigation.

The NF1 gene is very large, with coding regions, or exons, scattered over a wide region of genomic DNA. To date, several thousand base pairs of its sequence have been determined. The first discovery to emerge from the sequence data was a similarity between the predicted protein product of the NF1 gene and the catalytic region of certain proteins, called GTPase-activating proteins or GAPs, that are known to interact with growth regulators within the cell. Experiments showed that, in fact, the NF1 protein could sub-

stitute for GAPs in switching off growth signals that might otherwise result in uncontrolled proliferation of cells and produce a tumor.

Thus the NF1 product may have a role in human biology that is much broader than its relevance to neurofibromatosis, as important as that will prove to be for NF1 sufferers and their families. The NF1 protein is likely to open new avenues of inquiry about normal cellular processes and about events that lead to cancer. Projects to investigate several of these avenues are now under way in our laboratory.

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 College 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 and as Associate Professor of Biochemistry and Biophysics before attaining his present position. Dr. Wiley is Chairman of the Committee on Higher Degrees in Biophysics. He is a Fellow of the American Academy of Arts and Sciences and a recipient of the Louisa Gross Horwitz Prize from Columbia University. Dr. Wiley was recently elected to the National Academy of Sciences.

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. Each organism has only a small number of different histocompatibility molecules (probably less than a dozen), so that each membrane 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 few years our laboratory has determined the three-dimensional structure of two human-class histocompatibility antigens (HLA) by high-resolution x-ray crystallography. This resulted in the visualization of antigen presentation, when an obvious peptide-binding cleft was revealed and shown to be filled with a presumed mixture of peptide antigens. Comparison of the structures of two polymorphic alleles of the human histocompatibility antigen refined to 2.6 Å resolution also suggested that the basis for allelic specificity, or "restriction," is in a set of pockets in the peptide-binding cleft.

We recently crystallized one of the histocompatibility antigens complexed with a peptide from influenza virus. This required reconstituting HLA from its two polypeptide subunits in the presence of a "restricted" peptide. It appears that the HLA molecule required peptide to be reconstituted efficiently *in vitro*, arguing that peptide binding may be linked to subunit assembly or stability.

In the past year, in collaboration with Jack Strominger and Joan Gorga, we crystallized a number of class II histocompatibility antigens and the complexes formed with superantigens.

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. Our laboratory has synthesized a number of these new ligands to confirm an atomic model for virus-cell binding that we proposed two years ago.

This year, in collaboration with Jeremy Knowles, Gary Glick, and George Whitesides, we have analyzed the binding of virus to multidentate ligands and have discovered some dimeric and polymeric ligands with increased affinity. 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 receptor, CD4.

Studies of Blood Cell Formation

David A. Williams, M.D.—Assistant Investigator

Dr. Williams is also Assistant Professor of Pediatrics at Harvard Medical School. He received his undergraduate degree in biology from Indiana State University and his medical degree from Indiana University School of Medicine. 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.

THE general goal of our research is to gain a better understanding of blood cell formation and of how to apply the knowledge to the treatment of relevant diseases. The work in progress is roughly divided into research on the role of the bone marrow environment in maintaining blood-forming cells in a normal way, on the effect of directing the expression of growth-regulating genes in specific blood cells in murine models, and on the development of gene transfer as a potential method of correcting human diseases in which blood cells are affected.

Investigating the Bone Marrow Environment

Blood cells are produced and delivered to the circulation by precursor, or stem, cells in the bone marrow. Balance is carefully maintained between production and utilization. Although advances have recently been made in explaining the control of production of daughter cells, the production of the stem cells is still little understood.

Bone marrow stem cells reside in a complex environment comprising many specialized cells, called stromal cells. This microenvironment can be recapitulated *in vitro*. However, investigation of stem cell interactions with the microenvironment has been hampered in the past by the complex nature of the stromal components. Using molecular methods, our laboratory, in collaboration with Vikrum Patel at Northwestern University, has generated permanent cell lines derived from this stromal milieu, and some of these lines have been shown to maintain blood formation. Moreover, they have led to the identification of important proteins of the bone marrow that help to anchor primitive blood-forming cells in the appropriate locations. Interaction of blood-forming cells with such proteins may be important in a variety of normal and abnormal functions, including control of blood formation, regeneration of stem cells after marrow transplantation, and movement of leukemia cells out of the bone marrow cavity.

In another project, cell lines generated from mice exhibiting a severe genetic abnormality in the bone marrow microenvironment have helped to identify the causative molecular lesion as well as the normal protein that this defective gene should produce. Work done in collaboration with Kris Zsebo at Amgen has shown that the protein is an important growth factor for very primitive blood-forming cells. It may prove to be a valuable therapeutic tool for bone marrow diseases.

In addition, Steven Paul of our laboratory, in collaboration with Yu Chung Yang (currently at Indiana University) and Genetics Institute, has utilized a stromal cell line generated from primate bone marrow to identify and clone another new growth factor affecting blood-forming cells and lymphocytes. This factor, called interleukin-11 (IL-11), appears to stimulate the growth of B lymphocytes and the bone marrow cells that give rise to blood platelets. Further characterization of the effects of this growth factor in mice is under way.

Regulated Expression of Growth-regulating Genes in Hematopoietic Cells

In order to study the effects of certain growth-regulating genes on the behavior of hematopoietic stem cells, the laboratory is utilizing embryonic stem (ES) cells to generate transgenic mice containing foreign genes. Such genes are easily introduced into cultured cells. Upon transfer of the manipulated cells into embryos, ES cells contribute to all the tissues of the resulting animals. This new and powerful technology allows the expression of new genes in specific hematopoietic cells (such as red or white blood cells) in adult mice.

Deniz Toksoz of our laboratory is investigating the effects of expression of specific growth factor genes on the development of blood-forming cells in the early mouse embryo. The earliest stage of blood formation occurs in the fetal yolk sac. Growth-regulating genes, such as macrophage or granulocyte colony-stimulating factors, have well-defined pharmacologic activities on hemato-

poietic cells in adult animals. However, less is known about the role of these factors in normal functioning of the hematopoietic system during fetal development. Through use of ES cell lines that have been programmed to make large amounts of these growth factors during the fetal yolk sac phase of embryo development, new insights into the control of blood formation in the embryo may be gained.

Gene Transfer and Somatic Gene Therapy

For several years we have been studying the use of gene transfer to treat human genetic diseases. Slow but steady progress has been made, using viral vehicles (called vectors) to transfer a correcting gene into hematopoietic stem cells. The laboratory has studied the transfer of adenosine deaminase (ADA), a lack of which can lead to a fatal disease called severe combined immunodeficiency (SCID). Jane Apperley and Barry Luskey, in a continuing collaboration with Stuart Orkin (HHMI, at the Children's Hospital, Boston), has demonstrated that ADA can be transferred into hematopoietic stem cells. Once these cells are transplanted into irradiated mice, the human ADA gene expresses functional protein at nearly the levels of the endogenous mouse ADA gene. Such levels are sustained in 75 percent of the animals for longer than six months, demonstrating stable expression of the introduced gene after complete reconstitution of the hematopoietic system by the transplanted cells. This work demonstrates that the introduction and expression of a vector-borne gene appears to be a viable approach to the correction of some severe diseases of the hematopoietic system in humans.

In addition, Barry Luskey, in collaboration with Arthur Neinhuis and David Bodine at the National Institutes of Health, has begun studies aimed at transferring the ADA gene into monkey bone marrow cells. These studies are very important prior to any use of retroviral technology to treat human disease.

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 the Massachusetts General Hospital. 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.

DURING embryonic development, tissues grow by the rapid proliferation of their constituent cells. An important mechanism by which intricate patterns of embryonic cell growth are regulated involves the local production of protein growth factors that stimulate cell proliferation. In adults the cells of most tissues proliferate slowly, dividing at a minimal rate that is just sufficient to replace senescent cells. However, in a few specific situations (for example, after tissue injury) the rate of cell proliferation can increase dramatically. This exceptionally rapid cell division is similar to embryonic cell proliferation and appears to be stimulated by some of the same growth factors.

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

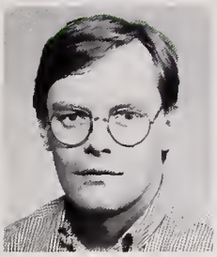
designed and produced mutant receptors that have normal PDGF-binding domains but defective kinase domains. When these mutants are introduced into cells that have normal receptors, a dimer (heterodimer) is formed between the normal and kinase-defective receptors. The normal and mutant receptors in the 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 receptors 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 identified one of these molecules, phosphatidylinositol-3 kinase (PI3 kinase), that appears to play an important role in PDGF-stimulated proliferation. Other investigators have found that PI3 kinase is also important in the cell transformation caused by some mammalian oncogenic viruses. Using the receptor as a probe, we recently purified the PI3 kinase and cloned the gene for this signaling molecule. We hope that by studying the interaction of the receptor and PI3 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 the setting of 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.

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 Assistant 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 and conducted postdoctoral research at the Massachusetts Institute of Technology, Whitehead Institute, with Richard Mulligan. His laboratory is investigating ways to treat genetic diseases by correcting the relevant genetic defects.

GENE replacement therapy is 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. A major focus of my laboratory is the development of new therapies for the treatment of cardiopulmonary diseases based on gene transfer into hepatocytes and airway epithelial cells.

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 plasma cholesterol. Hepatocytes are the appropriate target for gene transfer in this disease, because the liver is the primary organ 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 for this disease. 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 hepatocyte *in vitro*, and transplanting the genetically modified hepatocytes back into the affected recipient. We have used recombinant retroviruses to transfer a functional human LDL receptor gene stably into hepatocytes derived from WHHL rabbits. Transplantation of these cells into WHHL rabbits leads to substantial decreases in serum cholesterol for over six months.

An alternative and less-invasive approach is to deliver a functional LDL receptor gene into the hepatocytes *in vivo*. We have collaborated with George and Cathy Wu (University of Connecticut) to develop 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 high-level expression in the liver for at least four months following gene transfer. Administration of a gene transfer substrate capable of expressing LDL receptor 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 atherosclerotic disease in FH and other hyperlipidemic states.

Cystic fibrosis (CF), another inherited disease, is characterized by abnormal salt and water transport that leads to pathology within the pancreas and lungs. CF is the most common congenital disease among Caucasians, with a prevalence of 1 in 2,000 births. 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 patient with CF by introducing into them a normal CF gene. Replication-defective viruses were used to shuttle a normal CF gene into pancreatic cells of 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 of CF at a cellular level.

We have now entered the second and more difficult phase of the development of CF gene therapies, in which we are attempting to deliver normal CF genes to the relevant somatic cells of the body, the cells that line the airways of the lung. The overall strategy is to develop a gene transfer substrate capable of delivering a functional CF gene to airway cells *in vivo*. Direct inhalation of the gene transfer substrate could provide a noninvasive way for delivering the genes to the appropriate cell.

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. He holds 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 in molecular virology with David Baltimore at the Massachusetts Institute of Technology before joining the faculty at UCLA. Last year he was awarded the Milken Family Medical Foundation Award in Basic 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. Millions of B lymphocytes are produced daily in the fetal liver or adult bone marrow from primitive stem cells, and a careful balance between growth rate and development must be maintained for the stem cells and their progeny. 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 these stem cells and their production of B lymphocytes under normal circumstances and in various disease states.

The ABL Oncogene in Human and Murine 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, including those of immature cell types within the B cell lineage. The biological properties of the *ABL* gene product depend on its activity as a tyrosine-specific protein 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 called the Philadelphia chromosome or Ph1. This chromosome translocation uses messenger RNA splicing to join portions of chromosome 22 sequences encoding a gene called *BCR* to a portion of the coding sequences of the *ABL* gene on chromosome 9. The tyrosine kinase activity of the chimeric *BCR/ABL* gene product is evoked under these circumstances 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 cases of Ph1-positive acute lymphocytic leukemia, a P185 BCR/ABL protein product is commonly found. Surprisingly, the precise contribution of *BCR* sequences determines the efficiency of tyrosine kinase activity of the *ABL* segment and the malignant potential of the gene product.

Detailed site-directed mutagenesis studies on molecularly cloned copies of the *BCR/ABL* genes show that *BCR* sequences are essential to activate the *ABL* tyrosine kinase activity and produce a functional oncogene. The precise function of BCR in the normal cell is not known, but several lines of evidence, including sequence homologies, autophosphorylation activity, and inactivation with nucleotide analogues, suggest that BCR may be a protein kinase itself. We are evaluating the possibility that the enzymatic action of BCR is required to activate the enzyme function of ABL in the chimeric oncogene.

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* in other types of leukemias because the normal *ABL* gene can be toxic to most rodent fibroblast cell types when highly expressed via gene transfection techniques. The precise mechanism of toxicity is not established, but probably relates to a cell cycle blocking effect.

To circumvent this problem, we have used a new approach to preparing retroviral expression stocks developed by Dan Littman and his colleagues (HHMI, University of California, San Francisco). 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 produced 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*. We will further ana-

lyze such mutant strains for their biological potential and test for analogous mutations in naturally occurring human leukemias.

The activation of the cellular *ABL* gene to oncogenic forms correlates with the stable *in vivo* modification of the protein with phosphotyrosine. The normal ABL proteins are not detectably modified with phosphotyrosine *in vivo*, although they can autophosphorylate themselves *in vitro*. We have recently demonstrated that a specific intracellular inhibitor suppresses the tyrosine kinase activity of the normal cellular ABL proteins, but does not effect the oncogenic forms. If normal ABL proteins are hyperexpressed, they can overcome this inhibitor and modify themselves on tyrosine *in vivo*. In further studies on this inhibitor, we will try to define its mechanism of differential action on the normal and oncogenic forms of the ABL proteins.

Stem and Progenitor Cell Growth Regulation

One possible mechanism of action of oncogenic forms of *ABL* would be a transcriptional activation of specific growth factor genes, leading to an autocrine stimulation of receptors in different cell types. We have tested this idea in a pre–B cell line that is interleukin-7 (IL-7) dependent. Although all oncogenic forms of *ABL* could relieve the cell of growth factor dependence and lead to tumor progression, there were interesting differences between Abelson virus and the *BCR/ABL* forms. A subset of the lines derived with Abelson virus showed activation of a low level of transcription from the IL-7 locus, but none of the *BCR/ABL* lines showed this activation. When forced autocrine expression of IL-7 was engineered in the same indicator line, it could render the cells growth factor–independent, but the cells did not progress to tumorigenic behavior. It is likely that a bypass of the growth factor requirements and additional signals initiated by the oncogenic forms of *ABL* determine tumor progression for immature B lineage cell types.

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 like liver or kidney and show shorter latency than animals infected with the P210 *BCR/ABL* form of the gene. 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.

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* and, in preliminary analysis, also show evidence of reconstituting other lineages. We hope to develop clonal lines that will retain these functional properties.

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.

CELLS use diverse molecular mechanisms for regulating the synthesis of proteins in response to environmental changes. Any of the many steps in the pathway by which the information stored in genes becomes converted into proteins can be subject to regulation. My laboratory is particularly interested in understanding the process by which messenger RNA (mRNA) becomes translated into proteins. We wish to understand fundamental aspects of this process, as well as how translation of certain proteins can be regulated in response to conditions outside the cell.

The information in mRNA is translated into protein by a large RNA-protein complex, the ribosome. It has been known for some time that the movement of ribosomes along the mRNA is not linear with time. Rather, they pause at discrete places along the way. Why they do so is not well understood, but is thought to be due either to rare codons or to secondary structures in the mRNA.

As a postdoctoral fellow in Peter Walter's laboratory, I devised a new method to determine the distribution of ribosomes on an mRNA with single-nucleotide precision. Using this assay, we have been able to detect and map discrete sites of ribosome pausing during translation in extracts. Two of these rate-limiting steps correspond to initiation and termination of translation. Others occur during elongation of the nascent protein. We are using this method to identify features of mRNA sequence or structure that result in ribosome pausing.

Surprisingly, we have found that the pausing causes a sort of traffic jam, in that ribosomes pile up behind each paused one. Under certain conditions, we have detected up to nine ribosomes piled up behind the leading one. As a result, the ribosomes become tightly packed, such that their centers are only 27 nucleotides apart. This tight packing along the mRNA forces the ribosomes in each cluster to orient themselves in a zigzag pattern.

Although we have only detected ribosome stacking in cell extracts, we are interested in determining if this phenomenon occurs in intact cells. "Zigzag" or "helical" polyribosomes have been previously seen in the electron micrographs of certain cells, but their functional significance has been unclear. It seems likely that they represent densely stacked ribosome clusters, caused by the pausing of the leading ribosome.

We are also investigating the mechanism by which the translation of a secretory protein, insulin, is regulated in response to conditions outside the cell. Insulin is the major protein involved in regulating glucose metabolism in mammals. It is synthesized by the β cells of the pancreas, which are organized into clusters known as islets. When primary cultures of rat pancreatic islets are incubated in high concentrations of glucose, the synthesis of the insulin precursor molecule, proinsulin, increases between 5- and 10-fold within minutes. During the first hour of glucose administration, this increased synthesis of insulin is largely due to more efficient translation of preexisting mRNA. Although the phenomenon has been described for many years, little is known of the molecular mechanisms involved.

We are using the above assay, in conjunction with biochemical fractionation, to probe the mechanism by which the translation of proinsulin mRNA is regulated in pancreatic islets. Our initial results indicate that preproinsulin translation is regulated at both the initiation and elongation levels. We hope that a better understanding of this phenomenon will be helpful in explaining how, in certain patients with diabetes, normal insulin production is impaired.

Molecular Genetics and 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 of the Department of Psychiatry.

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.

Prenatal Diagnosis

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 can be used to trace the transmission of individual PAH genes from the parents to the children in PKU families. A fetus who has inherited 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 many western European countries.

Carrier Detection

The above procedure is applicable only to families with affected children. To prevent PKU from occurring in the first place, it is necessary to identify the asymptomatic carriers, which requires extensive knowledge of the gene defects. To determine the molecular basis of PKU among Caucasians, patient samples were obtained from various European countries and molecularly analyzed. The defects in the 10 most common mutant PAH genes and a number of less prevalent ones have been characterized.

The 10 prevalent mutations together represent about two-thirds of all PKU genes in Caucasians. If carrier screening is carried out today and followed by prenatal diagnosis in couples who are both carriers, it is theoretically possible to reduce the incidence of PKU by ninefold. Thus the number of PKU patients in this country could potentially be reduced from 300–400 a year to 27–36 a year. It is important to continue with the molecular analysis: if 90 percent of PKU genes could be detected throughout the Caucasian population, it would be possible to reduce the number of PKU patients to only 3–4 a year.

Population Dynamics

A number of these PKU genes have distinctive patterns of distribution on the European continent. One is very prevalent in eastern Europe, and the 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 the people in prehistoric times. Similarly, another PKU gene is very frequent in northern Europe but less frequent in the neighboring countries. These results suggested that multiple PKU mutations occurred independently in various parts of Europe and then spread throughout the Caucasian race.

Somatic Gene Therapy

The other major goal of our laboratory is to explore the potential for somatic gene therapy of the genetic disorder. The PAH gene has been inserted into the genome of an incapacitated virus. The recombinant viruses are able to infect mammalian cells and incorporate themselves into the genome of the infected cells, but they are no longer able to produce new virus to continue the infection process. The recombinant viruses were used to infect cultured rodent hepatoma cells, thereby conferring on them an ability to synthesize the corresponding human enzyme. More recently we have demonstrated that normal hepatic cells from mice can be cultured and efficiently infected by the recombinant virus.

These results have led to experimentation with hepatocyte transplantation in laboratory animals. A variety of inert substances were used as support for the mouse hepatic cells prior to transplantation. Unfortunately, the transplanted hepatocytes

lived for only a few weeks in the recipient animals. However, when the hepatocytes were returned to the liver of mice by direct injection into the portal vein or the spleen, the cells incorporated themselves into the liver parenchyma and lived for as long as the mice did. They also continued to function as hepatocytes in the transplanted animals. These exciting results should form the foundation for the curing of liver disorders in humans by gene therapy in the future.

The investigation of the feasibility of gene therapy for PKU has been hampered by the lack of an animal model. Recently, the laboratory of William Dove at the University of Wisconsin has reported the creation by chemical mutagenesis of a mouse that is deficient in PAH activity in the liver. Our initial experiments with these mice have shown that the PAH gene can indeed be transferred into their hepatocytes in culture and reconstitute fully the missing enzyme activity. This mouse model will be critically important for us to test the efficacy of our hepatic gene transfer and hepatocyte transplantation protocols for correction of the hepatic deficiency.

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 the 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 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.

We have now gone on to ask the complementary question—whether we can quantitatively predict the background adaptation in photoreceptors from the known properties of the Ca^{2+} feedback. First, we derived the time course of light-induced phosphodiesterase activity (which hydrolyzes cGMP) from a cell's response to a dim flash in the condition of no feedback. Next, we introduced the equations that describe the Ca^{2+} feedback, the components of which include a Ca^{2+} influx through the cGMP-gated conductance, a Ca^{2+} efflux through a Na^+-Ca^{2+} exchange mechanism, and the Ca^{2+} modulation on the guanylate cyclase. All of these processes are known in quantitative detail from previous work.

Computations showed that such a simple model of the Ca^{2+} feedback could indeed predict quite well a cell's physiological response to a dim flash. Furthermore, the model led to a steady-state response-intensity relation for steps of light that also fitted well the experimental relation. The predicted response-intensity relation is consistent with Weber's law, which constitutes the classical description of background light adaptation. In summary, there is both experimental and theoretical support for the Ca^{2+} feedback to be the predominant mechanism underlying background adaptation in photoreceptors.

Another problem we are working on is a molecular characterization of the cGMP-activated conductance mediating phototransduction. This conductance now appears to belong to a family of cyclic nucleotide–gated channels that includes a channel in olfactory cilia supposedly mediating olfactory transduction. The known channels in this family so far are all slightly different from one another in their functional characteristics. For example, the rod and cone channels show similar activation by cGMP but differ in the current-voltage relation. The olfactory channel, on the other hand, differs from both the rod and cone

channels by having a much higher affinity for cGMP (and also adenosine 3′:5′-cyclic monophosphate, or cAMP, another important cyclic nucleotide).

Recently the rod channel in bovine retina was purified and cloned by others. On the basis of structural homology to the rod channel, Randall Reed, another HHMI investigator at Johns Hopkins, has also obtained a cDNA clone for the rat olfactory channel. We examined the physiology of this channel, after expression in a mammalian cell line, and found that it has a current-voltage relation remarkably similar to that of the rod channel, except that cGMP and cAMP activate this channel about 50 times more effectively than they do the rod channel. The next step to be taken is to generate hybrids between the two channels, in order to identify the part of the molecules that determines affinities for cyclic nucleotides.

In addition, with Dr. Reed's help, we have isolated, again by structural homology, several clones from a human retinal cDNA library that may encode for both the rod and cone channels in the human retina. Experiments are now under way to test these clones for functional expression.

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 is using two complementary approaches to relate ion channel function to structure. 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 first strategy 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 this 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 signal transduction in this protein.

A second approach for finding parts of the protein that are important for specific functions involves mutant selection. We are developing a system in which we can randomly mutagenize the gene for a channel and then select the interesting mutants. This strategy is almost the reverse of site-directed mutagenesis. Instead of choosing to change a particular part of the protein and looking for the resultant changes in function, we select for a change in function and then look at the underlying structural changes. We have expressed the genes for the nicotinic acetylcholine receptor in tissue-cultured mammalian cell lines in preparation for this mutant selection strategy. We have also found a method for selecting certain channel mutants by using a fluorescently tagged toxin molecule that binds differently to these mutant channels.

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.*

OF the billions of cells composing the mammalian brain, a few thousand, located in the hypothalamus, function together to form a biological clock. This clock can control the timing of daily behaviors such as sleep and wakefulness with an accuracy of minutes. Although chemical and electrical rhythms have been detected in these mammalian neural pacemakers, little is known about the underlying biochemistry used by these cells to compute time.

In a simpler model organism, the fruit fly *Drosophila*, genes and proteins central to biological timing are beginning to be recovered and characterized. Mutations in several genes alter the fruit fly's circadian (daily) rhythms of locomotor activity, which can be compared to wake/sleep cycles in mammals. The same mutations also alter noncircadian, high-frequency rhythms. Normal fruit flies produce a courtship song composed of wing-beating cycles with a periodicity of about a minute. At the cellular level, changes in chemical and electrical conductances are measured in some of the mutants, possibly linking measurement of time to controlled transmission of signals between cells.

The best-studied gene in the *Drosophila* clock system is the *per* (*period*) locus. Three mutant forms of *per* have been recovered that affect the pace of the insect's clock. In the per^{l} mutant, circadian rhythms have a long period of 30 rather than 24 hours. For the mutant per^{s}, daily cycles have a shortened, 19-hour period. The per^{o} mutants have no daily rhythms. Corresponding changes are found in the courtship song; an 80-second song for per^{l}, 40 seconds for per^{s}, and song arrhythmicity for per^{o}. For at least some tissues, per^{o} and per^{l} mutants exhibit lower than normal levels of intercellular communication, while per^{s} mutants show hypernormal communication. These effects appear to be mediated by changes in conductance of specialized channels, or gap junctions, between cells.

The molecular changes associated with the mutations have been established. Although per^{o} cannot make a full-length protein, per^{s} and per^{l} each make a *per* protein that is altered by substitution of a different, single amino acid. Comparable changes in cycle time can also be produced by altering the amount of *per* protein the fly produces. For example, *Drosophila* that have received by microinjection a new gene that underproduces the *per* protein 20-fold have 40-hour daily rhythms.

A variety of experiments have shown us that the *per* protein functions in the nervous system to control daily and circaminuten rhythms. Recently we have become interested in tracking the development of the fly's clock, in an effort to identify the first cells making the *per* protein. A *Drosophila* embryo receiving a flash of light only hours after its conception will develop into an adult fly showing a daily routine of activity in phase with the timing of the embryonic light signal. The fact that an adult fly can "remember" this timing must mean that a clock was running when the light signal was received and has been running ever since. These experiments show that the embryo's circadian clock starts running within 12 hours of fertilization of the egg, a time at which as few as 50 cells may be found making the *per* protein.

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 other indispensable genes. Of special interest is the *soiree* gene. The *soiree* mutants are arrhythmic in constant darkness but show nocturnal behavior in the presence of a light/dark cycle. The per^{o} mutants are diurnal in the same light cycles. In behavioral tests of per^{o} *soiree* double mutants, the nocturnal behavior is expressed, indicating a more central role for the new gene.

Development of Skin, Muscle, and Nerve

The nervous system and skin are derived from a common set of cells in the embryo, the ectoderm. Each cell within the ectoderm must choose a fate, and in certain mutants of *Drosophila* the choice goes awry. Neurogenic mutants have lost the capacity to choose skin, so that only nerve is formed. We believe these mutants have lost the

ability to form a set of early acting developmental signals, so we are using the mutations 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. The gene produces a very large protein (2,700 amino acids) that is 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 suspect that the neighbors read that signal and send signals to the *Notch*-bearing cell in return to come up with the correct allocations of skin and nerve. Biochemical experiments have shown that the entire string of hormone copies is tethered to the cell's surface; thus signaling between cells must be intimate. Only cells that can touch each other could communicate through such a protein.

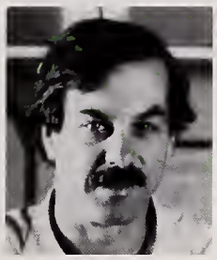
From work with temperature-sensitive mutations we know that *Notch* proteins are used to instruct development throughout embryonic, larval, and pupal life. Mutations altering the structures of individual hormone-like elements of the *Notch* protein have been examined to understand the role each plays in early and late development. Of the 36 hormone repeats, no two are identical and, importantly, changes in different hormone elements produce different developmental abnormalities. Thus alternate parts of the *Notch* hormone string must be 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 to form adult tissue.

What do signals from these genes tell a cell to do? For several years it has been argued that gene action at *Notch*, *Delta*, *big brain*, *almondex*, *neuralized*, *mastermind*, and *Enhancer of split* stimulates an undifferentiated ectodermal cell to develop as an epidermal cell; the genes are expected to provide a series of epidermalizing signals during cell differentiation, with loss of function generating a nerve cell. New work shows that this simple picture cannot be accurate.

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 can be read by cells composing different germ layers, with no apparent overlap in the cells' final developmental fates. 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. This last conclusion follows from our observation that 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.

Control of Transcription by Transmembrane Signals



Edward B. Ziff, Ph.D.—Investigator

Dr. Ziff is also Professor of Biochemistry at the 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 the 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 a tumor may arise. When the pathways are blocked, essential cell types may degenerate and die. It follows that errors in signaling can result in diverse sorts of 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.

We have shown that a protein kinase, Raf, which phosphorylates proteins on serine and threonine residues and which resides primarily in the cytoplasm, has a role in transmitting these activating signals to the c-*fos* promoter in the nucleus. Mutations that deregulate Raf can induce human tumors, and our work suggests that this may follow from aberrant activation of the expression of early response genes such as c-*fos*.

We have also identified novel transcription factor proteins that bind to regulatory regions of the c-*fos* promoter and activate the gene in response to second messenger signals. One factor is responsive to cAMP and moves from the cytoplasm to the nucleus upon activation. It apparently cooperates with other factors to enable a wide range of transmembrane signals to activate the gene.

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 other factors to induce the gene for tyrosine hydroxylase (TH), an enzyme that catalyzes a critical step in the production of neurotransmitters in the catecholamine family. 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.

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. Thus the mechanisms that control peripherin expression appear to be quite distinct from those that control c-Fos or TH.

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 uncontrolled expression of c-*myc*, another growth factor–induced gene. It encodes a protein, c-Myc, that is distantly related to c-Fos and has a 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 that can stimulate its DNA binding. Studies of the complexes of c-Myc with DNA may reveal the role of c-Myc in normal cell proliferation and in tumorigenesis.

We find that c-Myc expression is elevated in the naturally occurring childhood brain tumor medulloblastoma, where it is accompanied by a block to differentiation. 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 seek to understand how cells respond to growth factor signals by proliferating and how cells coordinate their final differentiation with their exit from the cell cycle. Future studies will emphasize the roles of transmembrane signaling in adult nervous system function.

Molecular Genetics of Visual Transduction

Charles S. Zuker, Ph.D.—Associate Investigator

Dr. Zuker is also Associate Professor of Biology and of Neuroscience 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, a McKnight Scholar, and a Fellow of the Alfred P. Sloan Foundation.

THE nervous system provides a dramatic example of the capabilities of cells to signal each other and to respond to environmental cues. An understanding of signal transduction mechanisms is essential for elucidating the cellular and molecular basis of information processing in biological systems. Phototransduction, the neuronal excitation process triggered by light and resulting in the generation of a receptor potential, provides an ideal model system in which to study signal transduction in response to an environmental stimulus. The aim of our research program is to elucidate mechanisms used for signal transduction in the visual system, using a combined molecular, genetic, and physiological approach.

Drosophila is unique as an experimental organism in that it allows a multidisciplinary approach to biological questions that will yield not only much new information, but may well provide a type of information not otherwise obtainable. First, the system is amenable to classical genetics: over the past decade several groups have isolated a large number of *Drosophila* mutants with specific defects in behavior mediated by visual input. Many of these have been shown to define genes important for phototransduction. Second, since the eye is fully dispensable, it can be manipulated without affecting viability. Third, genes can be readily introduced into the *Drosophila* germline. Thus genes and the proteins they encode can be experimentally manipulated *in vitro* and their functions studied *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. Recently our efforts have focused on the characterization of genes encoding products involved in the intermediate events of the visual transduction cascade.

An example of the power and utility of a genetic approach to identify molecules with unexpected roles is the *ninaA* gene. *ninaA* mutants have a 10- to 20-fold reduction of rhodopsin levels. Characterization of the *ninaA* gene has shown that it encodes a visual system–specific polypeptide with high–amino acid sequence identity to the human cyclosporin A–binding protein, cyclophilin. Cyclosporin A is an immunosuppressing drug widely used to prevent graft rejection following transplant surgery. Cyclophilins are peptidyl-prolyl cis-trans isomerases that catalyze isomerization about the peptide bond between a proline residue (Pro) and its amino neighbor (Xaa). A model for *ninaA* function is that opsin is a substrate for the *ninaA* protein; isomerization about one or several Xaa-Pro peptide bonds may be crucial for proper membrane intercalation, folding, or stability of opsins.

In addition to *ninaA*, we are also studying protein kinases and two arrestins involved in regulating this signaling process. We are using molecular, genetic, and physiological approaches to dissect the role of these proteins *in vivo*. The results of these studies should help us understand the molecular basis of sensory reception and information processing and may help us understand abnormalities in the human visual and nervous systems.

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.
Shapiro, Larry J., M.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., Ph.D.
Kobilka, Brian K., M.D.
Nusse, Roeland, 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

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.
Davis, C. Geoffrey, 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

Amara, Susan G., Ph.D.
Artavanis-Tsakonas, Spyridon, Ph.D.

Bothwell, Alfred L. M., Ph.D.
Bottomly, H. Kim, Ph.D.
Brünger, Axel T., Ph.D.
Flavell, Richard A., Ph.D.
Fox, Robert O., Ph.D.
Horwich, Arthur, M.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.
Wolin, Sandra L., M.D., Ph.D.

ILLINOIS

Chicago
The University of Chicago and The University of Chicago Hospitals

Bell, Graeme I., Ph.D.
Fuchs, Elaine V., 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.

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.
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.
Larsen, P. Reed, M.D.
Maas, Richard L., M.D., Ph.D.

The Children's Hospital

Kunkel, Louis M., Ph.D.
Nadal-Ginard, Bernardo, M.D., Ph.D.
Orkin, Stuart H., M.D.
Williams, David A., M.D.

Harvard Medical School and Brigham and Women's Hospital

Church, George M., 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.
Schwartz, Benjamin D., M.D., 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

Alt, Frederick W., Ph.D.
Axel, Richard, M.D.
Hendrickson, Wayne A., Ph.D.
Jessell, Thomas M., Ph.D.
Kandel, Eric R., M.D.
Schwartz, James H., M.D., Ph.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.
Cullen, Bryan R., Ph.D.
Green, Warner C., M.D., 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.
George, Donna L., Ph.D.
Kadesch, Thomas R., Ph.D.
Liebhaber, Stephen A., M.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.
Watterson, D. Martin, Ph.D.

TEXAS

Dallas
University of Texas Southwestern Medical Center at Dallas and associated hospitals

Beutler, Bruce A., M.D.
Deisenhofer, Johann, Ph.D.
Ellis, Leland, Ph.D.
Fischer Lindahl, Kirsten, Ph.D.
Garbers, David L., Ph.D.
Gething, Mary-Jane H., Ph.D.
Katz, Flora N., Ph.D.
Sprang, Stephen R., Ph.D.
Südhof, Thomas C., M.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.
Rich, Robert R., M.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 and University of Utah Medical Center

Capecchi, Mario R., Ph.D.
Gesteland, Raymond F., Ph.D.
Hazelrigg, Tulle I., 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.

Index