VOLUME II GENERAL PRINCIPLES

# MOLECULAR BIOLOGY OF THE GENE

FOURTH EDITION

James D. Watson

Nancy H. Hopkins

Jeffrey W. Roberts

Joan Argetsinger Steitz

Alan M. Weiner

VOLUME II SPECIALIZED ASPECTS

# MOLECULAR BIOLOGY OF THE GENE

FOURTH EDITION


**James D. Watson** COLD SPRING HARBOR LABORATORY

**Nancy H. Hopkins** MASSACHUSETTS INSTITUTE OF TECHNOLOGY

**Jeffrey W. Roberts** CORNELL UNIVERSITY

**Joan Argetsinger Steitz** YALE UNIVERSITY

**Alan M. Weiner** YALE UNIVERSITY



**The Benjamin/Cummings Publishing Company, Inc.**

Menlo Park, California • Reading, Massachusetts • Don Mills, Ontario
Wokingham, U.K. • Amsterdam • Sydney • Singapore
Tokyo • Madrid • Bogota • Santiago • San Juan


# About the Authors

**James D. Watson** is the Director of the Cold Spring Harbor Laboratory. He spent his undergraduate years at the University of Chicago and received his Ph.D. in 1950 from Indiana University. Between 1950 and 1953 he did postdoctoral research in Copenhagen and Cambridge, England. While at Cambridge, he began the collaboration that resulted in 1953 in the elucidation of the double-helical structure of DNA. (For this discovery, Watson, Francis Crick, and Maurice Wilkins were awarded the Nobel Prize in 1962.) Later in 1953 he went to the California Institute of Technology. He moved to Harvard in 1955, where he taught and did research on RNA synthesis and protein synthesis until 1976. While at Harvard he also wrote the first, second, and third editions of *Molecular Biology of the Gene*, which were published in 1965, 1970, and 1976, respectively. He has been at Cold Spring Harbor since 1968, where his major interest has been the induction of cancer by viruses.

**Nancy H. Hopkins** is a Professor of Biology at the Massachusetts Institute of Technology. She graduated from Radcliffe College in 1964 and did graduate work at Yale and Harvard, receiving her Ph.D. in Molecular Biology and Biochemistry from Harvard in 1971. After postdoctoral work at the Cold Spring Harbor Laboratory, she joined the faculty at M.I.T., where she teaches and does research on the molecular biology of retroviruses. She is the primary author of Chapters 23 through 27 in Volume II of this edition of *Molecular Biology of the Gene.*

**Jeffrey W. Roberts** is a Professor of Biochemistry at Cornell University. He received a B.A. in Physics and Liberal Arts from the University of Texas in 1964 and a Ph.D. in Biophysics from Harvard in 1970. He was a postdoctoral fellow at Harvard and also did research at the MRC Laboratory of Molecular Biology in Cambridge, England, before going to Cornell in 1974. His current research interests are genetic regulation in bacteria and phages, in particular the regulation of transcription and the control of DNA repair functions. He is the primary author of Chapters 11, 12, 13, 16, and 17 in Volume I of this text.

**Joan Argetsinger Steitz** is a Professor of Molecular Biophysics and Biochemistry at Yale University. She graduated from Antioch College in 1963 and received a Ph.D. from Harvard in 1967. She did postdoctoral work at the MRC Laboratory of Molecular Biology before joining the Yale faculty in 1970. Her research interests have always focused on the structures and functions of RNA molecules; her current research is on gene expression in mammalian cells, with an emphasis on the roles of small RNA-protein complexes. A member of the National Academy of Sciences, she is a recipient of the National Medal of Science, among other awards. She is the primary author of Chapters 14, 15, 20, and 21 in Volume I of this text.

**Alan M. Weiner** is a Professor of Molecular Biophysics and Biochemistry at Yale University. He graduated from Yale College in 1968 and received his Ph.D. from Harvard in 1973. After postdoctoral work at Stanford University and M.I.T., he returned to Yale as a faculty member in 1976. His current research concentrates on the structure, function, and evolution of mammalian genes for small nuclear RNA species. He is the primary author of Chapters 22 and 28 in Volume II of this text.

# Preface

Today no molecular biologist knows all the important facts about the gene. This was not the case in 1965 when the first edition of *Molecular Biology of the Gene* appeared. Then there were few practicing molecular biologists and not too many facts to learn. So what we knew about DNA and RNA could easily be explained to beginning college students. That year the final codons of the genetic code were being assigned, and everyone at the forefront of research could regularly assemble in the modest lecture hall at Cold Spring Harbor. Five years later, when the second edition appeared, our numbers were rising rapidly. Yet, despite the emerging popularity of molecular biology, it was still quite uncertain if the future would be as intellectually meaningful as the years just after the discovery of the double helix. The isolation of the first repressors and the demonstration that they bind specifically to control sequences in DNA seemed to some pioneers in DNA research to mark the end of the years of germinal discovery. With no means to isolate the genes of any higher organism, much less any way to know their nucleotide sequences, any pathway to understanding how genes guide the differentiation events that give rise to multicellular organisms seemed impossibly remote.

Happily, these worries did not last long. By the time the third edition of *Molecular Biology of the Gene* was published (1976), recombinant DNA procedures had given us the power to clone genes. Moreover, there was reason to believe that highly reliable methods to rapidly sequence long stretches of DNA would soon be available. As this new era of molecular biology began, however, there initially was widely voiced concern that recombinant DNA procedures might generate dangerous and pathogenic new organisms. It was not until after much deliberation that in 1977 the cloning of the genes of higher organisms began in earnest. The third edition could barely mention the potential of recombinant DNA, and of necessity its brief discussions of how genes function in eucaryotic organisms were tentative, and sometimes quite speculative.

It is only in this fourth edition that we see the extraordinary fruits of the recombinant DNA revolution. Hardly any contemporary experiment on gene structure or function is done today without recourse to ever more powerful methods for cloning and sequencing genes. As a result, we are barraged daily by arresting new facts of such importance that we seldom can relax long enough to take comfort in the accomplishments of the immediate past. The science described in this edition is by any measure an extraordinary example of human achievement.

Because of the immense breadth of today's research on the gene, none of us can speak with real authority except in those areas where our own research efforts are concentrated. Thus it was clear from the first discussions about the fourth edition that writing it would be beyond the capability of any one scientist who also had other major responsibilities. So the task of preparing this edition has required several authors. We also realized that it would be a formidable undertaking to keep the book within a manageable length; even by adopting a larger page format, we saw no way not to exceed a thousand pages. DNA can no longer be portrayed with the grandeur it deserves in a handy volume that would be pleasant to carry across a campus. Although this edition could have been shortened by eliminating the introductory material found in the first eight chapters, we never seriously considered this alternative. To do so would remove the background material that so many readers of previous editions have found valuable, and which has let many novices in molecular biology use this book as their first real introduction to gene structure and function.

Now that we are at last finished, we find that the book is even longer than we had planned. In part this happened because we are two years behind schedule, and

150 additional pages were needed to accommodate the immediate past. We also seriously underestimated how many words and illustrations would be required to describe the extraordinary variety of gene structures and functions that underlie the complexity of eucaryotic cells. We therefore have made the decision to split the fourth edition into two volumes. In the first volume we cover the general principles that govern the structure and function of both procaryotic and eucaryotic genes. It can be used as the sole text for a one-term course in molecular biology at the undergraduate level. The second volume concentrates on those specialized aspects of the gene that underlie multicellular existence, and it concludes with a chapter on the evolution of DNA. In this edition the second volume is appreciably smaller than the first. This will not be true of subsequent editions. Now that it is at last possible to study differentiation at the DNA level, we can easily foresee the time when, in fact, more than one volume will be required for even an introductory description of how genes are organized and expressed in the specialized cells of multicellular organisms.

We hope that this new edition, like its predecessors, will be found to be a highly suitable text for teaching at the undergraduate level, and that it also will provide all molecular biologists with an easy reference to the basic facts about genes. We have shown sections of the manuscripts to a variety of colleagues who are listed as reviewers. Their comments have been taken seriously, and we hope that the final manuscript faithfully reflects their expertise. Any mistakes that remain are, of course, our responsibility. Those who have made major contributions by writing or rewriting large sections of the text are Thomas Steitz (Chapter 6), Ira Herskowitz (Chapters 18 and 19), John Coffin (Chapter 24), and Brent Cochran (Chapter 25). Their generous contributions of specialized knowledge has vastly upgraded those portions of the book. In addition, John Coffin, Scott Powers, Haruo Saito, Lisa Steiner, and Parmjit Jat helped with the references for various chapters in Volume II. The excellent index was prepared by Maija Hinkle.

Equally important have been the efforts at Cold Spring Harbor of Andrea Stephenson, whose competent secretarial assistance helped coordinate our diverse labors, and Susan Scheib, whose intelligent attention to detail kept the manuscript and the galleys moving on a forward course. We also wish to acknowledge the pleasure of working with the staff of The Benjamin/Cummings Publishing Company, including Editor-in-Chief Jim Behnke and Production Supervisors Karen Gulliver and Betsy Dilernia. In particular we wish to thank Jane Gillen, who has functioned as the responsible editor during the entire writing and production of the book. An especially satisfying aspect of the process has been seeing rough drawings come alive through the efforts of the talented illustrator Georg Klatt, who has been responsible for the vast majority of the hundreds of new drawings prepared for this edition, and whose commitment and interest have greatly improved the book. And finally we gratefully acknowledge the strong support of our families throughout this endeavor, which was of course far more difficult and protracted than we ever foresaw.

James D. Watson

Nancy H. Hopkins

Jeffrey W. Roberts

Joan Argetsinger Steitz

Alan M. Weiner

# Reviewers

John Abelson, California Institute of Technology
Bruce Alberts, University of California-San Francisco
Manny Ares, Yale University
Spyros Artanvanis-Tsakonas, Yale University
Piet Borst, The Netherlands Cancer Institute
Clifford Brunk, University of California-Los Angeles
Tom Cech, University of Colorado
Brent Cochran, Massachusetts Institute of Technology
John Coffin, Tufts University
Nick Cozzarelli, University of California-Berkeley
Don Crothers, Yale University
Steve Dellaporta, Cold Spring Harbor Laboratory
Ashley Dunn, Ludwig Institute for Cancer Research, Melbourne
Gary Felsenfeld, National Institutes of Health
John Fessler, University of California-Los Angeles
Wally Gilbert, Harvard University
Nigel Godson, New York University
Howard Green, Harvard Medical School
Nigel Grindley, Yale University
Carol Gross, University of Wisconsin
Ron Guggenheimer, Cold Spring Harbor Laboratory
Gary Gussin, University of Iowa
David Helfman, Cold Spring Harbor Laboratory
Roger Hendrix, University of Pittsburgh
Ira Herskowitz, University of California-San Francisco
Andy Hiatt, Cold Spring Harbor Laboratory
Bob Horvitz, Massachusetts Institute of Technology
David Housman, Massachusetts Institute of Technology
Martha Howe, University of Wisconsin
Richard Hynes, Massachusetts Institute of Technology
Amar Klar, Cold Spring Harbor Laboratory
Nancy Kleckner, Harvard University
Larry Klobutcher, University of Connecticut
Chuck Kurland, Uppsala University
Peter Laird, The Netherlands Cancer Institute
Richard Losick, Harvard University
Brenda Lowe, Cold Spring Harbor Laboratory
Tony Mahowald, Case Western Reserve University
Nancy Maizels, Yale University
Will McClure, Carnegie Mellon University
Bill McGinnis, Yale University
Jeffrey Miller, University of California-Los Angeles
Paul Modrich, Duke University
Peter Moore, Yale University
Steve Mount, University of California-Berkeley
Kim Mowry, Yale University
Tim Nelson, Yale University
Harry Noller, University of California-Santa Cruz
Bill Nunn, University of California-Irvine
Paul Nurse, Imperial Cancer Research Fund
Kim Nasmyth, Laboratory of Molecular Biology, Cambridge
Norm Pace, Indiana University
Carl Parker, California Institute of Technology
Sheldon Penman, Massachusetts Institute of Technology
Scott Powers, Cold Spring Harbor Laboratory
Mark Ptashne, Harvard University
Charles Radding, Yale University
David Raulet, Massachusetts Institute of Technology
Phil Robbins, Massachusetts Institute of Technology
Hugh Robertson, Rockefeller University
Lucia Rothman-Denes, University of Chicago
Earl Ruley, Massachusetts Institute of Technology
Haruo Saito, Harvard University
Robert Schleif, Brandeis University
Thomas Shenk, Princeton University
Gerry Smith, Hutchinson Cancer Center, University of Washington
Deborah Steege, Duke University
Lisa Steiner, Massachusetts Institute of Technology
Tom Steitz, Yale University
Bruce Stillman, Cold Spring Harbor Laboratory
Bill Studier, Brookhaven National Laboratory
Bob Symons, University of Adelaide
Susumu Tonegawa, Massachusetts Institute of Technology
Olke Uhlenbeck, University of Colorado
Axel Ullrich, Genentech
Chris Walsh, Massachusetts Institute of Technology
Jim Wang, Harvard University
Bob Webster, Duke University
Allan Wilson, University of California-Berkeley
Barbara Wold, California Institute of Technology
Sandra Wolin, University of California-San Francisco
Keith Yamamoto, University of California-San Francisco
Jorge Yunis, University of Minnesota

# Brief Contents

# Detailed Contents

VOLUME II SPECIALIZED ASPECTS

CHAPTER 22

# The Molecular Biology of Development

**Embryology** began as a descriptive science, growing out of the fascination that biologists experienced as they watched the remarkable events of early development in organisms as diverse as the fruit fly, the sea urchin, and the mouse. For example, the female frog deposits her eggs on the bottom of a pond, and the male frog fertilizes them. Shortly thereafter, tiny tadpoles can be seen freely swimming about, searching for food and shelter, completely on their own in a very large (and often dangerous) world. Viewed under the microscope, both the huge, nearly spherical frog egg and the much smaller, flagellated sperm have a characteristic morphology, but neither of these germ cells resembles the tadpole or the adult frog in any obvious way. How, then, do these two highly specialized germ cells fuse and develop so quickly into an independent organism with a functional digestive tract and a nervous system capable of coordinating rapid swimming motions? We know that the instructions for how the egg develops into an adult are written in the linear sequence of bases along the DNA of the germ cells. However, this genetic information would be useless if the fertilized egg could not express the information in an orderly fashion. During development, gene expression must therefore be regulated both in space (an adult fly must not have a leg in the middle of its forehead) and in time (the larval fly must not prematurely develop wings). In addition, it is not sufficient for the DNA within each cell to be properly expressed; the cells must also interact with one another so as to build complicated multicellular structures such as wings and legs. Embryologists search for the general principles of development by concentrating on the early developmental stages in the life of the organism when cellular differentiation and interaction are most apparent.

## The Heart of Embryology Is the Problem of Cell Differentiation

All higher plants and animals are constructed from a large variety of cell types (e.g., nerve cells, muscle cells, thyroid cells, and blood cells) that must arise in an exquisitely coordinated way. In some organisms, specialization begins with the first few cell divisions after fertilization. In other organisms, a large number of divisions occur before any progeny cell is fixed in its fate. Regardless of the exact time that **differentiation** occurs, however, it always results in the transformation of the parental cell into a large number of morphologically different progeny cell types.

Classical developmental biologists examined differentiation from three viewpoints. First, what are the external (and internal) influences acting on the original undifferentiated cell that might initiate a chain of events resulting in two progeny cells of different constitution? Sometimes, asymmetrically acting external forces are easy to perceive. For example, gravity forces the yolk of an amphibian egg to the bottom. Thus, after the first few cell divisions subdivide the fertilized egg, some of the embryonic cells have more yolk than others.

Second, are the molecular differences between differentiated cells extreme, or does morphological differentiation arise from the presence of only a few unique proteins in especially large numbers? We now know that each type of differentiated cell contains many molecular species peculiar to that cell type. Thus, a complete description of differentiation at the molecular level would necessarily be a most formidable task, and would not necessarily help us to understand the essential mechanisms responsible for differentiation.

Third, are the various changes that bring about differentiation irreversible, and if so, how they are perpetuated in a heritable fashion from one cell to the next? As we shall see, these are not easy questions to answer. Whether differentiation is reversible or irreversible depends not only on which cell type we are talking about, and in which organism, but also to a surprising extent on the precise details of the experiment. Moreover, we are only just now beginning to have vague hints of the mechanisms by which cells can maintain or change their state of differentiation.

Until quite recently, embryology was largely studied as an isolated subject, apart from modern genetic or biochemical ideas. Now, however, it is clear that the morphological tools of the classical embryologist cannot give satisfying answers by themselves. Instead, as in genetics, fundamental answers require analysis at the molecular level. Thus, just as recent methodological advances have made certain aspects of biochemistry and genetics indistinguishable, so embryology is being transformed by progress in biochemistry and genetics brought about by the recombinant DNA revolution (Chapter 19).

## A Hierarchy of Genes Controls Development[1, 2, 3]

How, then, can we understand development at the molecular level when differentiation involves so many changes in the protein composition of the cell (together with parallel changes in key metabolic pathways and intermediates, as well as in RNA, lipid, polysaccharide, and perhaps even ionic composition)? If each of the many hundreds of gene products that distinguish a liver cell from a muscle cell had to be studied in minute detail before we could begin to understand the molecular causes of development, the situation would be virtually hopeless. However, we shall see that for many, and perhaps all, developing organisms (both procaryotic and eucaryotic), the gene products responsible for development can be arranged in a hierarchy, with some genes controlling the expression of other genes. This conclusion should not really surprise us. For example, we have already seen that the λ CI protein (repressor) controls the developmental decision to grow lytically or to lysogenize (see Figure 17-16). Similarly, the *MAT* locus in yeast controls the expression of a large array of genes responsible for the differences between the two mating types and the diploid (see Figure 18-31). These are only a few obvious examples. Thus, the

many hundreds of gene products that distinguish one mammalian cell from another are not all equal; some are more important than others, because they control the initial decisions to differentiate along one developmental pathway or another. The goal of molecular biologists studying development is therefore to discover which genes control the expression of other genes.

Traditionally, one of the problems that has confounded and confused the molecular study of development is that cells usually become "committed" or "determined" to differentiate long before any actual morphological differentiation is apparent. For example, stem cells in the bone marrow of mammals give rise to at least two different kinds of progenitor cells—those that proliferate and differentiate into oxygen-carrying red blood cells and those that proliferate and differentiate into antibody-producing lymphocytes. Yet, in the very early stages of proliferation, the two kinds of progenitor cells (although committed to different fates) are difficult to distinguish. Once we realize that genes are arranged in a hierarchy, it becomes clear that in most cases, **commitment** or **determination** corresponds to the expression of a controlling gene, while differentiation reflects the myriad molecular consequences of that initial developmental decision.

## Necessity of Finding Good Model Systems for Studying Differentiation

We have already noted (Chapter 20) that the relative amount of genomic DNA increases by a factor of approximately 800 from *E. coli* to a mammalian cell, but not all of this additional DNA represents a real increase in genetic complexity. Satellite sequences, moderately repeated DNA sequence families, and introns account for much of the genome in higher cells. Nonetheless, any particular mammalian cell type (say, liver) does contain a much larger number of distinct protein species than does *E. coli*. In addition, although all of the several hundred different cell types in the body share a core of "housekeeping" proteins that are necessary for metabolism and replication, different mammalian cell types (say, liver and muscle) express very different subsets of the total protein-coding capacity of the genome.

Faced with such biological complexity, how are we to choose the best organism for studying selected aspects of differentiation? Historically, rapid progress in understanding *E. coli* was due primarily to the invention of powerful genetic techniques for dissecting complicated biochemical events; these experimental advantages, in turn, attracted the concentrated efforts of many scientists. Similarly, with only a few important exceptions, molecular biologists have tended to concentrate on organisms where relatively powerful genetic techniques can be brought to bear on otherwise intractable developmental problems. Thus, even if we are ultimately curious about human biology, common sense may direct us to work on a simpler or more convenient organism such as bakers' yeast *(Saccharomyces cerevisiae)*, the fruit fly *(Drosophila melanogaster)*, the soil nematode worm *(Caenorhabditis elegans)*, or the laboratory mouse *(Mus musculus)*. However, it is important to keep in mind that biologists have always been experimental opportunists, studying whatever systems promised to yield interesting results most readily, and today's molecular biologists are no different. Some organisms, such as the African clawed toad *(Xenopus laevis)* and the cellular slime mold *(Dictyostelium discoideum)*, have