# Understanding DNA and Gene Cloning

## *A Guide for the Curious*

*Karl Drlica*

# Understanding DNA and Gene Cloning
## *A Guide for the Curious*

*Karl Drlica*
*Public Health Research Institute*
*and New York University*

**John Wiley & Sons, Inc.**
*New York • Chichester • Brisbane • Toronto • Singapore*

*Acquisitions Editor* Sally Cheney
*Copy Editing Manager* Deborah Herbert
*Production Manager* Linda Muriello
*Senior Production Supervisor* Savoula Amanatidis
*Designer* Pete Noa
*Cover Design* Ben Arrington
*Illustration Coordinator* Sigmund Malinowdki
*Manufacturing Manager* Lorraine Fumoso



*Library of Congress Cataloging in Publication Data:*

Drlica, Karl.

Understanding DNA and gene cloning : a guide for the curious / Karl Drlica.—[2nd ed.]

p. cm.

Includes index.

ISBN 0-471-62225-7 (paper)

1. Molecular cloning. 2. Recombinant DNA. 3. Genetic engineering. I. Title.

QH442.2.D75 1992

574.87'328—dc20 91-26076

CIP

*Printed in the United States of America*

10 9 8 7 6 5 4 3 2

Printed and bound by Courier Companies, Inc.

# PREFACE

Gene cloning technologies continue to spur advances in many biological disciplines, and an update of *Understanding DNA* is long overdue. As with the first edition, my goal is to explain the fundamental principles of DNA biology at a level that does not require a knowledge of chemistry. Thus the discussion starts at a much more elementary level than is commonly found in publications such as *Scientific American* and *The New York Times*. It then builds on the basics to introduce the reader to some of the more sophisticated concepts of molecular biology by using drawings and in some cases analogies.

*Understanding DNA* was written for college students. In junior colleges the readers have often been potential biology majors, whereas in elite private universities they have tended to be nonscience majors. But it has been gratifying to see *Understanding DNA* reach a much wider audience, one that has included investment bankers, mechanical engineers, medicinal chemists, and precocious high school students. The book even caused a Ph.D. physiologist to begin using gene cloning in a research project involving kidney function.

Several elements have been added or expanded to make the second edition a more comprehensive teaching tool. One is the addition of questions for discussion. Some of these questions have specific answers to reinforce a point; others are open-ended to stimulate additional reading. Many of the questions also introduce information that would dilute the main themes if this information were included in the primary text. Another teaching aid is the glossary, which has been expanded. Vocabulary is a key aspect of learning molecular biology, and the reader should expect to refer frequently to the glossary. As an in-

structor I found that administering simple glossary quizzes early in a course overcomes some of the vocabulary barriers. A third aid is the list of additional readings, which has also been expanded. Most of the new entries are from *Scientific American* because the articles are of high quality, they are at the appropriate level when used in conjunction with *Understanding DNA*, and *Scientific American* is readily available.

Several important topics have been added to the second edition. Among them is a discussion of retroviruses, viruses that cause cancer and AIDS. AIDS is a relatively new disease that is affecting us all; even those not exposed to the virus itself are seeing personal relationships become more circumspect, are being troubled by controversial educational programs, and are witnessing medical institutions and insurance companies alter their services to cope with the expense of the disease. Another addition is a brief description of monoclonal antibodies, immunological tools that greatly expand our ability to study the biology of specific proteins. When combined with gene cloning strategies, monoclonal antibody techniques provide incredibly precise methods for learning how biological information is stored, transmitted, and used by cells. Other topics added are gene amplification (polymerase chain reaction), Southern hybridization, ribozymes, and the Ti plasmid.

I would like to thank a number of people for helping with the second edition. John Balbalas deserves special acknowledgement for the illustrations. Deborah Everett, Barry Kreiswirth, Ellen Murphy, Abraham Pinter, Steve Projan, Todd Steck, and Shermaine Tilley provided many excellent ideas and valuable criticisms; John Kornblum and Brenda Griffing painstakingly edited the manuscript and turned up many flaws that had escaped me in the first edition.

*Karl Drlica*

# PREFACE TO THE FIRST EDITION

An explosion of knowledge is shaking the science of biology, an explosion that will soon touch the life of each one of us. At its center is chemical information—information that our cells use, store, and pass on to subsequent generations. With this new knowledge comes the ability to manipulate chemical information, the ability to restructure the molecules that program living cells. Already this new technology is being used to solve problems in diverse areas such as waste disposal, synthesis of drugs, treatment of cancer, plant breeding, and diagnosis of human diseases. The new biology is also telling us how the chemicals in our bodies function; we may soon be programming ourselves and writing our own biological future. When this happens, each of us will be confronted with a new set of personal and political choices. Some of these difficult and controversial decisions are already upon us, and the choices will not get easier. Informed decisions require an understanding of molecular biology and recombinant DNA technology; this book is intended to provide that understanding.

Molecular biology is a science of complex ideas supported by test tube experiments with molecules. Consequently, the science has remained largely inaccessible to those without a knowledge of chemistry. I hope to change that situation—this book requires the reader to have little or no background in chemistry. Chemical processes and molecular structures are described by means of analogies using terms familiar to nonscientists. Technical terms have been kept to a minimum; where they must be introduced, they are accompanied by defi-

nitions. In addition, a glossary has been provided for easy reference; items in the glossary are in boldface type the first time they appear in the text.

It is also my intent to provide a sense of how informational molecules are manipulated experimentally. Integration of these details should help remove the mystery from gene cloning and expose the elegance and simplicity of the technology. I hope that this brief introduction to gene cloning will help you enjoy and appreciate the science of molecular biology for the art form that it is.

A number of people have helped me in this endeavor, and they deserve most of the credit for making this book readable. I especially thank Lynne Angerer, Betty Bonham, Tom Caraco, Cheryl Cicero, Lisa Dimitsopulos, Dianne Drlica, Karen Drlica, Rob Franco, Claire Gavin, Ed Goldstein, Brenda Griffing, George Hoch, Hiroko Holtfretter, Johannes Holtfretter, Lasse Lindahl, Stephen Manes, Bill Muchmore, Pat Pattison, Donna Riley, Peter Rowley, Ron Smith, Franklin Stahl, Todd Steck, Ilene Wagner, William Wasserman, Grace Wever, Bill Wishart, and Janice Zengel. I also acknowledge Alvin J. Clark and Henry Sobel for technical information and Ron Sapolsky for artistic insights used in early versions of the manuscript. Fred Corey and his staff at John Wiley & Sons provided excellent editorial assistance. The illustrations are the creative work of John Balbalis; where appropriate he has attempted to provide a sense of relative scale among the elements involved.

*Karl Drlica*

# INTRODUCTION

These days, all of us are constantly exposed to the on-going revolution in biological knowledge. One's daily newspaper is likely to contain an announcement of an advance in our understanding of a disease such as cancer and AIDS, or an article about how the rapid increase in our detailed knowledge of thousands of human genes is leading to new forms of disease diagnosis and therapy. Yet I have been told by members of the press that most of their readers know almost nothing about modern biology — and that most of them are even unclear about the difference between a chromosome, a gene, and a DNA molecule. If so, then something is drastically wrong.

As was pointed out more than sixty years ago, "the key to every biological problem must finally be sought in the cell, for every living organism is, or at sometime has been, a cell" (E. B. Wilson, "The Cell in Development and Heredity," 1925). Each of us, for example, originates as a single cell (formed from the fusion of one sperm cell and one egg cell), which grows and divides until it produces a highly organized cooperative of more than 10 million million cells — the adult human. A typical cell is so small that it would take 10,000 of them to cover the head of a pin. But, the relatively simple behaviors of individual cells have, in aggregate, a surprising power to explain even sophisticated properties of multicellular organisms — such as the memory stored in

---

This introduction is an extension of ideas originally expressed by Bruce Alberts in *The American Zoologist*, 1989.

the nerve networks in our brains, or the growth and patterning of a developing embryo. Those of us who understand nothing at all about cells, like those of us who know nothing at all of poetry, are missing out on one of the great pleasures of human existence — the search to understand ourselves.

But what exactly is a cell? Originally defined as the smallest unit of an organism that is alive, to biologists today a cell is nothing more (or, much better, nothing less) than a special collection of complex molecules, enclosed by a membrane and having the very special ability to reproduce itself from the much simpler molecules available in its surroundings. Speaking as a chemist, a cell is a self-replicating collection of catalysts — most of which are proteins. We know how this works in principle, but a complete understanding of a living cell will require that we know every reaction that occurs in it, so we can see how each component contributes to the self-replication of the entire unit. In time, this knowledge will come — although even the simplest known cell, the tiny bacterium known as *Mycoplasma*, is estimated to contain a total of 40,000 protein molecules, of about 600 different kinds.

Cells and organisms are very complex. But, because they have evolved to this complexity by a repeated process of DNA sequence duplication and divergence, each cell is composed of parts that are closely related to other parts of the same cell in their structure and function. This fact greatly simplifies the task of understanding the tens of thousands of proteins that make a human being. For the same reason, there is a surprising uniformity among living things. We know from DNA sequence analyses that plants and higher animals are closely related, not only to each other, but to relatively simple single-celled organisms such as yeasts. Cells are so similar in their structure and function that many of their proteins can be interchanged from one organism to another. For example, yeast cells share with human cells many of the central molecules that regulate their cell cycle, and several of the human proteins will substitute in the yeast cell for their yeast equivalents!

Scientists who have devoted their lives to studying cells view the cell as a large and elegant puzzle. Each biological macromolecule (protein, nucleic acid, or polysacharide) that is discovered and studied in detail represents a small piece of the puzzle, which will only be satisfactorily understood when it has been adequately connected to all of the other "pieces" in the cell with which it interacts. Ten years ago, our

total amount of information about cells was so small that most of the interconnections between the known pieces were missing. In the last few years, we seem to have reached the point where enough of the puzzle has been filled in that each new piece analyzed (most often a new protein) can often be connected to several others to provide some new insight. In terms of the puzzle analogy, we are still far from seeing the final picture, but we can often glimpse part of a tree, or recognize a familiar face in an otherwise chaotic jumble of partial information about the cell.

Much of the present excitement in biology stems from the feeling that we are now starting to know enough about cells to derive the type of connections that make conceptual sense of what seemed previously to be only an inexplicable muddle of facts. Moreover, by detailed comparisons of the components in the cells of *different* present-day organisms, we can hope to solve an even bigger puzzle than that of the workings of the cells themselves: what is the exact pathway by which living organisms evolved on the earth? This "megapuzzle" represents perhaps the ultimate intellectual challenge for future biologists.

The modern emphasis on explaining biological phenomena in terms of the behavior of molecules reflects the belief of today's biologists that the tools are in hand to achieve such a detailed, mechanistic explanation. This book describes the most revolutionary of these tools: gene cloning and the accompanying recombinant DNA technology. These methods, unforeseen as little as 20 years ago, have made it possible to answer almost any question about the cell, given sufficient effort. How quickly biology has changed! When I worked as a graduate student in the early 1960s, the cell seemed incomprehensibly complex. Most importantly, there seemed to be no obvious way of deciphering this complexity. Most of the tens of thousands of different protein molecules in a higher eukaryotic cell were known to be present in such small amounts that it appeared impossible to ever know their structure. As lucidly explained in this book, this situation has entirely changed. With gene cloning and the ability to manipulate the cloned genes so as to produce any gene product, every protein in the cell is potentially accessible in virtually unlimited amounts. Like the first settlers to arrive in California for the Gold Rush of 1849, today's biologist faces an easy harvest of riches. For the next twenty years or so, one need not be especially clever or wise to make a major contribution to biology. With luck, even the random cloning of a new gene—which

requires relatively little skill and no insight—can turn out to be exciting.

That cells exist at all is a marvel. To speak about them as "simply a self-replicating collection of catalysts" in no way reduces the beauty and wonder of the living state. If Earth were to be visited by a being from outer space, this being would undoubtedly find even the simplest of the living cells far more fascinating than any human-made object. That our children are largely bored with cells and the rest of biology—but fascinated by consumer electronics and automobiles—is a great tragedy, and it reflects how far we have to go in changing how science is taught to the general public. If this book by Karl Drlica can make a contribution to bringing an appreciation of the beauty of cells to others, the world will be much richer for it.

*Bruce M. Alberts*
*American Cancer Research Professor of Biochemistry*
*University of California, San Francisco*

# INTRODUCTION TO THE FIRST EDITION

In thinking about the course of human events, it has often occurred to me that they very much resemble the course of a river. A river meanders, gathers small streams, widens, deepens, and may even split into smaller rivers that go their separate ways. On occasion, rivers merge, a confluence that creates a mightier river. In the same sense, the extraordinary developments in genetic chemistry are part of an even more profound development in medical science, a change that is truly revolutionary. It is the confluence of the many discrete and previously unrelated medical science subjects into a single, unified discipline. Anatomy, physiology, biochemistry, microbiology, immunology, and genetics have now merged and are expressed in a common language of chemistry. By reducing structures and systems to molecular terms, all aspects of body form and function blend into a logical framework. Universities still maintain departmental lines to define administrative boundaries, but they are now meaningless in the pursuit of new knowledge.

The remarkable confluence of medical science first appeared in the genius of Louis Pasteur. More than any individual or school, he established medicine as a science and gave it the form we recognize today. Pasteur was trained as a chemist. His first exploit as a very young man was to show that two samples of tartaric acid of identical chemical composition differed physically because the molecules were mirror images of each other. Pasteur's "germ theory of disease" bore the stamp of his chemical background. He tried to reduce a problem of

disease to elementary components. His experimental approach was to purify the causative agents to homogeneity and recreate the disease with the isolated pure form of the agent. From this he created and practiced the disciplines of microbiology and immunology. It might surprise many microbiologists and immunologists today to find that in 1911 the *Encylopaedia Britannica* described Pasteur as a French *chemist*, the acknowledged head of the greatest *chemical* movement of his time.

Pasteur's scientific career had a flaw. Having established that the yeast cell is responsible for the conversion of sugar to alcohol, he tried to extract from the yeast cell the juices that would do the same thing. In this he failed and so concluded that nothing short of a living cell could possibly carry out this very complex chemical reaction. Pasteur's self-confidence, persuasiveness, and influence were so great that attempts by others to obtain alcoholic fermentation in a cell-free system were severely discouraged. And so, cellular vitalism became firmly rooted, and the advent of modern biochemistry was delayed for 30 years.

Only at the turn of this century did Eduard Büchner in Munich accidentally discover fermentation by disrupted yeast cells. In employing sugar as a preservative for yeast extracts, he observed a strange frothing. He had the insight to identify carbon dioxide as the gas and ethanol as the product of sugar degradation by the yeast juice. It was Pasteur's poor fortune that his extracts of Parisian yeast were deficient in sucrase, the enzyme that initiates the pathway of sugar metabolism. Luckily for Büchner, adequate amounts of the enzyme survived in his extracts from Munich yeast.

The reactions by which a yeast cell converts sugar to ethanol and carbon dioxide could then be isolated and analyzed in detail. In all, a dozen discrete, complex molecular rearrangements, condensations, and scissions are needed to achieve the fermentation of sugar to alcohol. Each of these reactions is catalyzed by an elaborate protein, an enzyme, designed to carry out the singular operation. The enzyme increases the rate of the reaction by a million- or trillionfold and gives it a unique direction among the many potential fates to which the molecule is susceptible.

These revelations of alcoholic fermentation in yeast provided the methods and confidence for the investigation of a comparable question. How does a muscle derive energy from sugar to do its work? When that mystery was unraveled, the plot and most of the characters

in the muscle story incredibly proved to be the same as in yeast. There is, of course, one deviation. In muscle at the final stage, lactic acid is produced instead of alcohol and carbon dioxide.

Reconstitution in the test tube of the yeast and muscle pathways of sugar combustion to generate usable energy set the stage for a generation of enzyme hunters in the 1940s and 1950s. My own attempts at synthesizing DNA with enzymes in a test tube were regarded by some as audacious. Reconstitution of the metabolism of fats as well as carbohydrates may be one thing, but the enzymatic synthesis of genetically precise DNA, thousands of times larger, must be quite another. Yet all I have done is follow in the classical traditions of biochemists of this century. It always seemed to me that a biochemist with a devotion to enzymes could, with sufficient effort, reconstitute in the test tube any metabolic event as well as the cell does it. In fact, the biochemist, freed from the cellular restraints of the concentrations of enzymes, substrates, ions, and metals, and with the license to introduce reagents that retard or drive a reaction, should do it even better.

As the disciplines of genetics, microbiology, and physiology reached more and more for chemical explanations, they began to coalesce with the biochemistry of the enzyme hunters. From this coalescence came molecular biology and genetic engineering. Narrowing our focus to the molecular biology of DNA, I would cite several diverse origins. One origin is in medical science. In 1944 Oswald Avery, in his lifelong and relentless search for control of pneumococcal pneumonia, became the first to show that DNA is the molecule in which genetic information is stored. A second origin of molecular biology is in microbial genetics. In the late 1940s and early 1950s microbiologists, some of them renegade physicists, chose the biology of the small bacterial viruses, the bacteriophages, to elucidate the functions of the major biomolecules: DNA, RNA, and proteins. At about the same time a third origin of molecular biology arose as the structural chemistry of these biomolecules became highly refined. Analysis of the X-ray diffraction patterns of proteins revealed their three-dimensional structures; the DNA patterns gave us the double helix and a major insight into its replication and function. A fourth origin of molecular biology is in biochemistry, the enzymology, analysis, and synthesis of nucleic acids. The biochemist provided access to the nucleases that cut and dissassemble DNA into its genes and constituent building blocks, the polymerases that reassemble them, and the ligases that link DNA

chains into genes and the genes into chromosomes; these are the reagents that have made genetic engineering possible. In the cell, these enzyme actions replicate, repair, and rearrange the genes and chromosomes.

Molecular biologists practice chemistry without calling it such. They identify and isolate genes from huge chromosomes, often only one part in millions or billions, and then they amplify that part by even larger magnitudes using microbial cloning procedures. They map human chromosomes, analyze their composition, isolate their components, redesign their genetic arrangement, and produce them in bacterial factories on a massive industrial scale. New species are created at will. Not even the boldest among us dreamed of this chemistry 10 years ago. I generally underestimated how permissive *E. coli* would be at accepting and expressing foreign genes. As the effects of a more profound grasp of chromosome structure and function become manifest, the impact on medicine and industry will prove to be far greater even than extrapolations from the current successes in the mass production by genetic engineering of rare hormones, vaccines, interferons, and enzymes.

Since the role of basic research is not always apparent to the general public, I would like to make another historical comment. Genetic engineering is solely an outgrowth of basic research. It was never planned, nor was it even clearly anticipated. Many of the procedures were discovered as unanticipated by-products of experiments designed to satisfy someone's curiosity about nature. For example, the analyses and rearrangements of DNA that form the drama of genetic engineering depend largely on a select cast of enzymes. Yet these actors were neither discovered nor created to fill these roles. Some of these enzymes, uncovered in my laboratory, came from a curiosity about the mechanisms of DNA replication. In these explorations, sponsored by the National Institutes of Health and the National Science Foundation for more than 25 years at a total cost of several million dollars, I neither anticipated nor promised their industrial application. Nor did any of my colleagues with comparable, federally funded projects. Thus, the multibillion dollar industry projected by Wall Street is entirely a product of the knowledge and opportunities gained from the pursuit of "irrelevant," basic research in universities, research made possible by the investment of many hundreds of millions of dollars by federal agencies over more than two decades.

As we retrace the flow of knowledge, we see that the first two decades of twentieth-century medical science were dominated by the microbe hunters. Their place in the spotlight was superseded for two decades by the vitamin hunters. They in turn were succeeded by the enzyme hunters in the 1940s and 1950s. For the past two decades the gene hunters have been in fashion. To whom the remaining years of our century will belong is uncertain. The neurobiologists, call them the head hunters, may very well claim it. If so, we will again see how chemistry is the fundamental language. Although brain chemistry may be novel and very complex, it is expressed in the familiar elements of carbon, nitrogen, oxygen, and hydrogen, of phosphorus and sulfur that constitute the rest of the body. Brain cells have the same DNA that all cells do; the basic enzyme patterns are those found elsewhere in the body. It is now known that hormones once thought to be unique to the brain are produced in the gut, ovary, and other tissues. The form and function of the brain and nervous system must ultimately be explained in terms of chemistry. The repeated failures of science to analyze social, economic, and political systems should not discourage us from pursuing the idea that individual human behavior, at least, can be explained by physical laws.

I sense in the future a better awareness that life can be described in rational terms and a furtherance of chemical language to express it. For chemistry is a truly international language. It links the physical and biological sciences, the atmospheric and earth sciences, the medical and agricultural sciences. The chemical language is a rich and fascinating language that creates images of great aesthetic beauty. I see the language of chemistry taught and used for the clearest statements about our individual selves, our environment, and our society. Such visions excite me. I hope you share them. They give us courage to face the future.

*Arthur Kornberg*
*Stanford University*

# CONTENTS