Understanding
DNA *and Gene*

CLONING

A Guide for the Curious

KARL DRLICA

Understanding **DNA** *and Gene*

CLONING

A Guide for the Curious

KARL DRLICA
University of Rochester

JOHN WILEY & SONS
New York • Chichester • Brisbane • Toronto • Singapore

Library of Congress Cataloging in Publication Data:

Drlica, Karl.
Understanding DNA and gene cloning.

Includes index.
1. Molecular cloning. 2. Genetic engineering.
I. Title. II. Title: Understanding D.N.A. and gene cloning.
[DNLM: 1. DNA, Recombinant. 2. Cloning, Molecular.
3. Genetic Intervention. 4. Molecular biology.
QH 442 D782u]
QH442.2.D75 1984 574.87'328 84-3518
ISBN 0-471-87942-8

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

PREFACE

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 definitions. 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

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 immu-

nology. It might surprise many microbiologists and immunologists today to find that in 1911 the *Encyclopaedia 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.

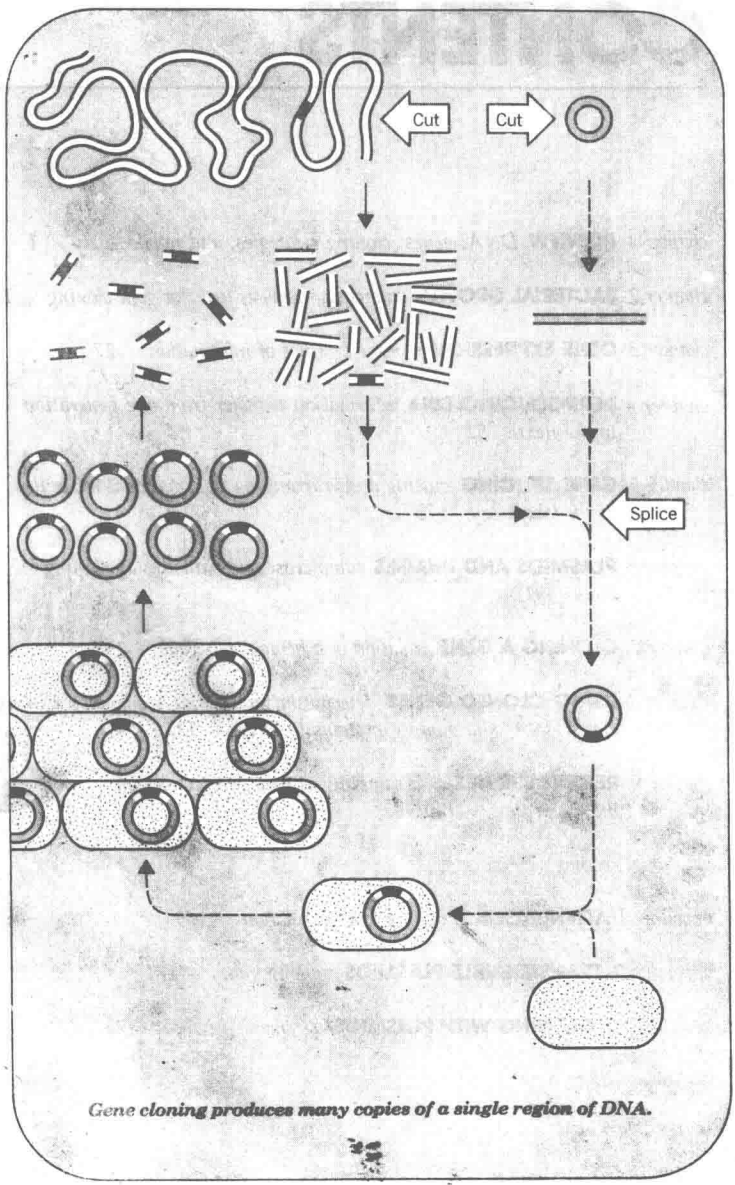
As we retrace the flow of knowledge, we see that the first two decades of twentieth-century medical science were dominated by the microbe hunters. Their place in the spotlight was superseded for two decades by the vitamin hunters. They in turn were succeeded by the enzyme hunters in the 1940s and 1950s. 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



Gene cloning produces many copies of a single region of DNA.

CONTENTS

chapter 1

PREVIEW

DNA, genes, cloning strategies, and public safety

overview

Information governing the characteristics of all organisms is stored in long, thin molecules of deoxyribonucleic acid (DNA). DNA molecules are divided into regions called genes, and genes control specific aspects of cellular chemistry. Methods are now available that allow biologists to cut any DNA molecule into specific fragments and to transfer individual fragments of DNA into a bacterium, a unicellular microorganism. A single bacterial cell receiving a specific fragment of DNA will divide repeatedly and form a cluster comprised of millions of identical cells. Each cell in a particular cluster, or clone, will contain many copies of the same DNA fragment. Billions of identical cells can be grown from a clone, and biologists can extract from these cells large amounts of a small region of DNA. That DNA is used for further study or possibly someday for correction of genetic diseases.

Since microorganisms receiving DNA fragments are being changed in unknown ways, gene cloning experiments have been viewed as potentially *dangerous. Regulations were therefore established to minimize the potential*

health hazards. Many types of recombinant DNA have now been constructed and studied; no harmful effects have been observed, and most molecular biologists now consider cloning of genes using nontransmissible plasmids to be safe (see Appendix II).

INTRODUCTION

In a general sense biologists have solved the riddle of heredity, the question of why offspring resemble their parents. We can now explain how heredity works by describing the chemical behavior of submicroscopic structures called molecules. At the center of this new understanding is a giant molecule called deoxyribonucleic acid (DNA). This book is about DNA, the chemical that specifies features such as eye color and blood type. DNA influences all our physical characteristics as well as those of every living organism on earth. This book is also about genetic engineering, recombinant DNA, and gene cloning. In particular, it is about how gene cloning works and what has been learned from it.

To begin, **cells**, **DNA**, and **gene cloning** must be described. (All boldface words are included in the glossary.) We can think of a cell as a self-reproducing bag of chemicals and microscopic structures. Our bodies are collections of trillions of cells working together. Each cell has its own identity and function. For example, liver cells cluster together to form livers, and skin cells attach to each other to cover our bodies. It is important to note that every cell contains all the components required for an independent existence; under the right conditions it is possible to isolate a human cell and get it to grow and divide in a test tube.

DNA is one of the components required for independent existence. It is a long, thin fiber that stores the information necessary to control the chemistry of life. DNA fibers are found

in every cell of our bodies, and they dictate how a particular cell behaves. Thus, DNA controls our body chemistry by controlling the chemistry of each of our cells.

Isolated DNA looks like a tangled mass of string (Figure 1-1);

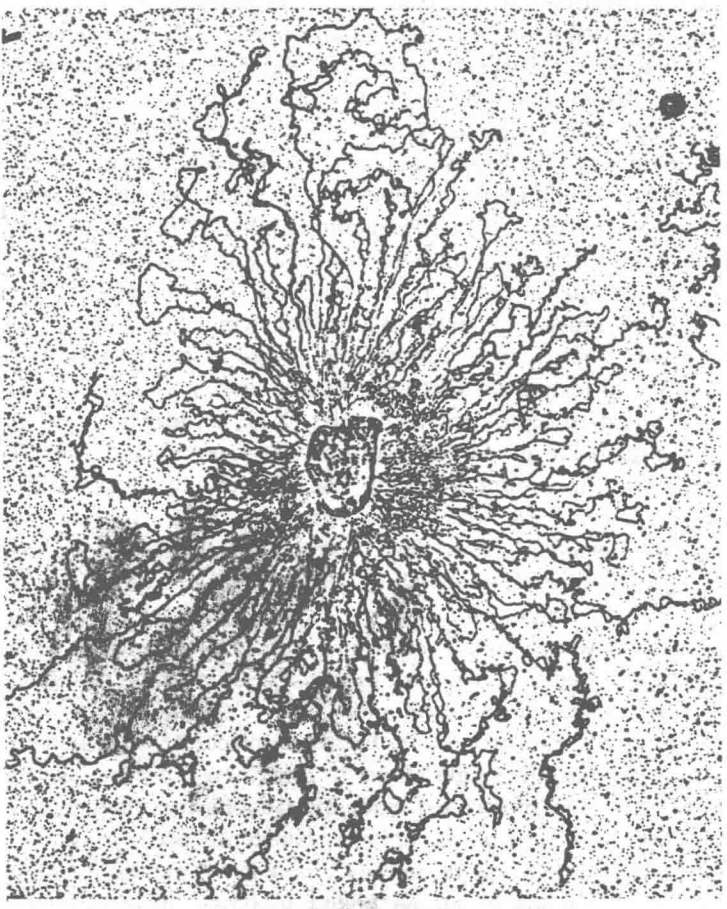


Figure 1-1. Electron Micrograph of a Single DNA Molecule Released from a Bacterium. *The long threadlike material is DNA and the dark mass in the center is the remnant of the bacterial cell wall. The entire DNA molecule is about one* ***millimeter*** *long. Photograph reprinted from Chromosoma by permission of Ruth Kavenoff, Oliver Ryder, and Springer-Verlag.*

each of our cells, which are generally less than a **millimeter** long, contains about two meters of DNA. Chemical analyses allow many detailed statements to be made about the structure of DNA, and in Chapters 3 and 4 some of these details are described. One of the goals of this book is to explain how this cellular information is stored, used, and reproduced. For now, the important concept is that a long, fibrous DNA molecule is subdivided into specific stretches or regions called **genes**. Each gene is responsible for causing a certain component to be made in the cell. The components interact to produce what we call life. Sometimes we can easily see the effects of particular genes; for example, a small group of genes is responsible for determining eye color. It is the specific information in the DNA, in the genes, that makes humans different from honey bees or fir trees. Information in your DNA makes you different from anyone else on earth—unless you have an identical twin.

Each DNA fiber is a **molecule**, a group of **atoms** joined together to form a distinct unit. Four points are important for understanding atoms and molecules. First, all forms of matter are composed of atoms. For example, water, wood, and steel are formed by the joining of **submicroscopic** building blocks called atoms. Second, the number of different kinds of atoms, or elements, is small (about 100). Atoms differ in size and in how they join with each other. Some are very reactive while others are inert. Hydrogen, for example, is very reactive; the airship *Hindenburg* would have been much safer if a nonreactive atom like helium had been used for lift. Third, only certain combinations of atoms join to form molecules. Thus molecules have discrete sizes and properties. Fourth, molecules can join with each other or with single atoms to form new combinations of atoms (i.e., new molecules). Such interactions are called chemical reactions. Chemical reactions are constantly occurring inside our cells. In chemical terms DNA is a giant molecule, often containing billions of atoms linked together. Information is stored in DNA by the specific arrangement of atoms, and this arrangement controls the chemical reactions in each of our cells.

Gene cloning developed in the mid-1970s when it became