BASIC METHODS IN

Molecular Biology

Leonard G. Davis, Ph.D.

Mark D. Dibner, Ph.D.

James F. Battey, M.D., Ph.D.

BASIC METHODS IN
Molecular Biology

Leonard G. Davis, Ph.D.
Neurobiology Group
Central Research and Development Department
E.I. Du Pont Experimental Station
Wilmington, Delaware

Mark D. Dibner, Ph.D.
Director, Business Studies Program
North Carolina Biotechnology Center
Research Triangle Park, North Carolina

James F. Battey, M.D., Ph.D.
National Cancer Institute
Medical Oncology Branch
National Naval Medical Center
Bethesda, Maryland

Elsevier
New York • Amsterdam • London

Cover Illustrations:

Representations of 24 base pairs of the standard "B" form of DNA, photographed on an Evans and Sutherland PS300 (Arnott, S., and Hukins, D., Biochem. Biophys. Res. Comm. *47*:1504, 1972). The molecular surface is displayed with dots (Connolly, M. L., Science *221*:709, 1983). Color coding is by atom type: nitrogen is blue; carbon is green; oxygen is red; phosphorus is yellow. The back cover shows the same molecule, cross-sectioned approximately halfway through the helix. Cover illustrations were created by and are courtesy of Dr. J. M. Blaney of the Biomedical Products Department of E. I. du Pont de Nemours and Company, Wilmington, Delaware.

Elsevier Science Publishing Co., Inc.
52 Vanderbilt Avenue, New York, New York 10017

Sole distributors outside the U.S.A. and Canada:

Elsevier Science Publishers B.V.
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Library of Congress Cataloging in Publication Data

Davis, Leonard G.
Methods in molecular biology.

Includes index.
1. Molecular biology—Methodology. I. Dibner, Mark D. II. Battey, James F. III. Title.
QH506.D39 1986 574.8'8'028 86-6316
ISBN 0-444-01082-3

Current printing (last digit):
10 9 8 7 6 5 4 3 2

Manufactured in the United States of America

Foreword

The heart of the most recent revolution in biology has been the development of the technology of genetics. Its achievements have simply changed what biologists do and, perhaps even more important, the way they think. Moreover, never before have scientists from such a broad range of disciplines rushed into such a small and slightly arcane field (as molecular geneticists used to believe theirs was) to learn, to carry off a bit of the technology, and to do it quickly because, armed with these powerful tools there was so much to do, so much to be learned. Doctors Davis, Dibner and Battey have done us a great service in providing the most powerful tool of all—an up-to-date, accessible, laboratory-tested, and comprehensive embodiment of what one needs to know to get on with the job at hand. They are experienced scientists. They state the principles and give the details. The rest is up to us.

Philip Leder
Boston

Acknowledgments

The authors wish to thank the many people and institutions that assisted us in compiling this text.

A number of scientists made valuable contributions to this text and deserve due recognition. Dr. Shoshana Segal provided the sections on mammalian cell transfection. Ms. Marian Kelley was the source of technical detail on monoclonal antibody preparation. Dr. Hunt Potter provided valuable help toward our writing of the section on electroporation. Ms. Donna Reed was a valuable resource for the preparation of protein methods. Dr. Eric Sinn contributed a large portion of our section on transgenic mouse analysis, and Dr. S. Carl Falco provided information for our comments on the use of yeast. Dr. Edward Sausville assisted in the preparation of the sections on cDNA cloning. The cover illustration was provided by Dr. J. Blaney.

Many other people provided assistance in the review of our manuscript. They include Drs. Philip Leder, Marion Cohen, Edward Berger, Rick Woychik, Keith Lawrence, Eric Seifter, Anne-Marie Lebacq, and Michael Kuehl. Sections of this manual were reviewed by Drs. J. Angulo, R. Arentzen, M. Lewis, F. Baldino, Jr., and L. Hennighausen. Additional help in reviewing was provided by A. Callahan, R. Manning, and R. Lampe.

We thank the people of Elsevier Science Publishing, and especially Mr. Yale Altman and Mr. Jonathan Wiener for their help and support in compiling this, our first book. Mr. Brian Trench created many of the drawings herein. We also thank New England Bio Labs, Bethesda Research Labs, Bio-Rad, Boehringer Mannheim, and Pharmacia for permission to reproduce selected illustrations. We thank Ms. Sherry M. Vari for her excellent word processing skills and long hours spent on text preparation.

Most of all, the authors are indebted to Dr. Philip Leder, Chairman of the Department of Genetics of the Harvard Medical School, for providing the leadership and resources responsible for the generation and evolution of most of these methods. We also thank past and present members of his laboratory group for sharing their knowledge with us to make this book possible.

Lastly, we wish to dedicate this book to our wives, Penny, Elaine, and Fran, who were wonderfully supportive of this project and the time it required.

Leonard G. Davis
Mark D. Dibner
James F. Battey
—April 1986

Contents

SECTION

1

The Basics of Molecular Biology

The Basics of Molecular Biology

Current biological science has been revolutionized by a series of new investigative techniques developed within the last 15 years. These techniques allow the definition of molecular mechanisms and structures that are responsible for such complex processes as cell growth and division, metabolism, differentiation and development. More significantly, they provide a way to manipulate molecules critical to these processes, and observe the changes in living systems that incorporate the altered molecules.

Nucleic acids and proteins are macromolecules; linear polymers comprised of subunits. Nucleic acids encode the genetic information specifying the primary structure of all proteins unique to an organism. Together with lipids and extracellular supporting stroma, they create cellular activity and physiological function. Thus, biological functions can be understood in part by examining the interrelationships between these key components. The genetic material of the cell, *deoxyribonucleic acid* (DNA), is a polymer composed of four nucleotide building blocks. Each of the four nucleotides contains a nucleic acid base (A, adenine; G, guanine; T, thymine; C, cytosine), a deoxyribose sugar moiety, and a phosphoester. Each strand is a string of nucleotides covalently bound together by phosphoester linkages between the 5′ carbon on the deoxyribose sugar of one nucleotide and the 3′ carbon of the sugar moiety on the neighboring nucleotide. Chains of these DNA subunits exist as two antiparallel strands in opposite polarity with respect to the phosphate sugar backbone, wound around each other in a double helical structure. One strand binds tightly to the other strand because there is the potential for hydrogen bond formation between specific b. es on one strand with bases on the opposite, or complementary, strand. Adenine is always paired with thymine, and guanine with cytosine. The fidelity of base pairing is provided by the nucleic acid synthesizing machinery that normally adds only the "correct" base specified by the template strand when elongating a new strand. It is the constancy and specificity of this complementary base pairing that forms the basis of DNA's function as a repository of genetic information. The order of nucleotides in DNA corresponds to the order of amino acids in proteins. As such, DNA can encode for proteins, with triplet groups of three adjacent nucleotides representing an mRNA codon, which specifies a particular amino acid. Therefore, the linear nucleotide sequence in DNA

specifies the order of amino acids for the cell's structural, functional, and enzymatic proteins. Other regions of DNA, which do not directly encode protein, contain information directing the regulation of gene product synthesis.

In the synthetic pathway between DNA and protein are the ribonucleic acids (RNA). The strand encoding the protein sequence information of the double-stranded DNA is copied, or transcribed, into a complementary strand of RNA. This RNA contains the same bases as DNA, except that uridine (U) is substituted for T and a ribose moiety is present instead of the deoxyribose. The RNA copy of the gene, called messenger RNA (mRNA), is translated with the assistance of transfer RNA (tRNA) and ribosomes (rRNA and associated proteins) to assemble sequentially the amino acids that form the primary sequence of protein.

Many molecular biology laboratory methods take advantage of the relative simplicity of prokaryotic cell systems such as bacteria. In prokaryotes, the continuous linear DNA sequence corresponds directly to linear RNA and protein sequences. However, in eukaryotes, the DNA encoding for protein cannot be read continuously as it contains interruptions (introns) in the translatable sequence. Eukaryotic DNA is thus first copied to a primary transcript (heteronuclear RNA) that is processed in the nucleus by excision of the protein coding sequences (exons). The exons are joined linearly into mature mRNA that can be processed further in the nucleus and moved to the cytoplasm for translation into protein. Certain newer methods allow the study of genes in eukaryotic cell systems.

Understanding the structure, function, and regulation of genes and their products is essential to an appreciation of biological systems. This also involves understanding the organization of an organism's nucleic acids. Previously this understanding was confounded by the complexity of the genome in eukaryotic cells, which contains up to 10^9 nucleotides in 50,000 genes. To analyze the genetic structure and events in this complex situation, one needs the ability to isolate and study a single gene in a purified form. Molecular cloning of DNA provides a mechanism for isolating a single discrete segment of DNA from a population of genes, purifying this segment to homogeneity, and amplifying the DNA segment to produce enough pure material for chemical, genetic, and biological analysis. The process of cloning relies entirely on performing enzymatic reactions in the laboratory, using well characterized bacterial DNA cleaving enzymes (restriction enzymes, REs) and modifying enzymes to copy, cut, and splice together discrete DNA molecules. DNA molecules are thus introduced into bacterial cells after being spliced into autonomously replicating DNA circles (plasmids) or bacterial viruses (bacteriophages). After many rounds of replication, the hybrid molecules are reisolated and purified, yielding sufficient quantities of the cloned DNA segment.

With the isolated, purified DNA segment the nucleotide sequence of bases can rapidly be determined, leading to the prediction of the amino acid sequence of the encoded protein. Radioactive labeling of this purified DNA allows the scientist to specifically probe for copies of related DNA sequences in complex cell genomes or related intracellular mRNA, amidst a background of up to a million unrelated sequences. mRNA synthesis from the purified DNA can be detected and quantitated in amounts as low as one to ten copies per cell.

Reengineering of the cloned DNA in bacteria or yeast may allow expression of its protein coding sequence, providing an inexpensive and abundant source of otherwise unattainable proteins of biological or medical importance. Alternative versions of the cloned DNA can be created in the laboratory by changing the structure or sequence. These DNA constructs can then be reintroduced into cells or whole animals to study the results of these man-made changes or mutations, and understand more completely the function and regulation of genes.

In this book, we describe methods for performing these experiments in molecular genetics. In each case, the method is described in a step-by-step, "cookbook" format and has been used, as written, with favorable results.

A WALK THROUGH THIS MANUAL

The methods in this book range from very simple to very complex. First is a description of the plasmid and vector systems and bacterial host cells used in the methods. The initial sections assume that a specific synthetic or cloned DNA probe is already available, allowing the selection, amplification, and examination of the gene of interest. Methods for isolating DNA from tissue, cutting the DNA to usable size, and separating the DNA pieces by size are discussed in Section 5. Sections 6 and 7 present methods for making probes, either synthetic or plasmid derived, to use in selecting DNA of interest. Methods for plasmid preparation and amplification are presented in Section 8. From the amplified plasmids, cloned DNA is excised and purified (Sections 9 and 10).

Section 11 turns to RNA—its preparation, selection, separation, and analysis. In Section 12, another type of cloning vector, the bacteriophage, is described. Please note that up to this point, the methods described involve the selection and amplification of DNA sequences that have already been cloned. The next two sections, 13 and 14, present methods for creating genomic DNA and cDNA libraries in bacteriophage vectors.

From the created library a desired clone is selected. The next step is to grow that DNA on a large scale, as described in Section 15 on subcloning into plasmids for preparative growth. From the higher yield of this cloned DNA, the sequence and other properties can be studied, following cloning into an appropriate M13 vector (Sections 16 and 17). Up to this point, DNA has been studied using the benefits of simpler prokaryotic systems. However, it may be of interest to put modified versions of the cloned gene back into the genome of eukaryotic cells in order to evaluate its regulation and function in a more biologically relevant system. Section 18 describes methods for incorporating DNA into mammalian cells growing in culture.

As mentioned above, proteins are the product of the genetic material, and it may be important to study them in order to understand gene regulation. Also, it is possible to translate RNA into proteins in vitro. These protein-related methods are described in Section 19.

The section on general methods (20) describes basic techniques that are incorporated into many of the other methods discussed in the text, such as DNA extraction, autoradiography, and titration of plaques. It is anticipated that the

novice will refer to these methods initially; in time they will become second nature.

Lastly, several more specialized molecular biological methods are described in Section 21. The first, transgenic mouse analysis, involves incorporation of new DNA pieces into a mouse embryo for later analysis in the postpartum animal. We also describe monoclonal antibody production techniques used to prepare immunological probes for specific gene products, as well as in situ hybridization, which uses nucleotide probes to localize and study specific genetic messages in tissue sections. Finally, some general notes are given on the use of yeast host and vector systems to perform molecular biology techniques.

The next few pages describe the use of specific techniques in molecular biological studies, with attention to questions that can be addressed using these methods.

SECTION

2

The Tools of the Molecular Biologist

The Tools of the Molecular Biologist

To illustrate the use of molecular biology methods, this section follows one possible series of experiments to study a typical gene, *X*, employing a variety of these methods.

It may be desirable to study gene *X* for its interesting structure or relevant expression in some biological context. Initially, a radiolabeled DNA probe needs to be obtained with a sequence similar to that on gene *X*, for example the gene from another species (homology to gene *X*). This probe can be purified and nick-translated to form a radiolabeled probe in order to detect the presence of gene *X* in a Southern blot analysis. Alternatively, a synthetic oligonucleotide probe can be synthesized in the laboratory to contain a sequence complementary to a portion of gene *X*. The labeled probe can then be used in DNA blotting to analyze DNA from a tissue or cell line of choice using DNA blots to define the presence of gene *X*–related sequences in the genome.

To do these DNA (Southern) blots, DNA from a tissue or cell line is isolated and purified and cut with specific restriction endonuclease(s) (REs) into defined fragments; the fragments of DNA are then fractionated by size using agarose gel electrophoresis. The DNA on the gel is transferred to a nitrocellulose filter (Southern blot), and the blot is hybridized with probe specific for gene *X* (Southern hybridization). The probe forms complementary base pairs only with restriction fragments that contain homologous sequences. Nonspecific radioactivity is washed away, and autoradiography of the blot demonstrates one or more bands if gene *X* is present or no bands if gene *X* is not found in the DNA tested.

An altered pattern of hybridizing DNA restriction fragments may appear on the Southern blot from DNA made from a specific tissue sample, indicating a change in the gene *X* structural sequences. For example, if there is a rearrangement of DNA in a specific tissue or tumor, this "somatic" rearrangement can be identified by purifying DNA from different tissue sources and probing, as described above. Genomic DNA from different cell types or tissues might show different size hybridizing fragments on the Southern blot, resulting from the changes introduced by rearrangement in the DNA.

Another example of an altered DNA pattern might be due to restriction fragment length polymorphisms (RFLPs) or different gene forms (alleles). If the genomic DNA from 100 individuals was cut with the RE *Eco*RI and was probed