NUCLEIC ACID ANALYSIS: PRINCIPLES & BIOAPPLICATIONS

NUCLEIC ACID ANALYSIS

PRINCIPLES AND BIOAPPLICATIONS

Editor

CHARLES A. DANGLER
Division of Comparative Medicine
Massachusetts Institute of Technology
Cambridge, Massachusetts

A JOHN WILEY & SONS, INC., PUBLICATION
New York • Chichester • Brisbane • Toronto • Singapore

Address All Inquiries to the Publisher
Wiley-Liss, Inc., 605 Third Avenue, New York, NY 10158-0012

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Nucleic acid analysis : principles and bioapplications / edited by
Charles A. Dangler.
p. cm.
Includes index.
ISBN 0-471-10358-6 (cloth : alk. paper)
1. Nucleic acid probes. 2. Nucleic acid probes—Diagnostic use.
I. Dangler, Charles A., 1957–
QP624.5.D73N82 1996
591.8′8—dc20 95-47202
CIP

The text of this book is printed on acid-free paper.

10 9 8 7 6 5 4 3 2 1

NUCLEIC ACID ANALYSIS

CONTRIBUTORS

Asim K. Bej, Department of Biology, University of Alabama at Birmingham, Birmingham, AL 35294-1170 **[1,231]**

Bruce Budowle, Forensic Science Research and Training Center, FBI Academy, Quantico, VA 22135 **[79]**

Jarasvech Chinsangaram, Department of Veterinary Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California, Davis, CA 95616 **[131]**

Catherine Theisen Comey, Forensic Science Research and Training Center, FBI Academy, Quantico, VA 22135 **[79]**

Charles A. Dangler, Department of Veterinary Science, The Pennsylvania State University, University Park, PA 16802; present address: Division of Comparative Medicine, Massachusetts Institute of Technology, Cambridge, MA 02139 **[31]**

Chitrita Debroy, Supelco, Inc., Supelco Park, Bellefonte, PA 16823 **[31]**

Peter Feng, Center for Food Safety and Applied Nutrition, FDA, Washington, DC 20204 **[203]**

Walter E. Hill, Seafood Products Research Center, FDA, Bothell, WA 98041-3012 **[203]**

Keith A. Lampel, Center for Food Safety and Applied Nutrition, FDA, Washington, DC 20204 **[203]**

James R. Lupski, Department of Molecular Human Genetics and Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030 **[157]**

Meena H. Mahbubani, Miles College, Birmingham, AL 35208 **[231]**

Edward R. B. McCabe, Department of Pediatrics, University of California, Los Angeles, CA 90024-1752 **[67]**

The numbers in brackets are the opening page numbers of the contributors' articles.

Linda L. McCabe, Department of Pediatrics, University of California, Los Angeles, CA 90024-1752 **[67]**

Bennie I. Osburn, Department of Veterinary Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California, Davis, CA 95616 **[131]**

Calvin H. Vary, Maine Medical Center Research Institute, South Portland, ME 04106 **[47]**

James Versalovic, Department of Molecular Human Genetics, Baylor College of Medicine, Houston, TX 77030 **[157]**

William C. Wilson, Arthropod-Borne Animal Diseases Research Laboratory, USDA, Agricultural Research Service, Laramie, WY 82071-3965 **[105]**

PREFACE

This text is intended for two types of readers. The first group is the relatively uninitiated, those persons not familiar with molecular biology, but who find themselves drawn into the world of nucleic acid technology because of the more applied scientific and commercial aspects evident on the horizon. Within this first group, scientists not already familiar with nucleic acid interactions and the development of nucleic acid assays will hopefully find this text valuable in helping them to become familiar with these topics. The text describes several approaches for generating nucleic acid probes and offers some advice on selecting labeling systems and assay formats. The second audience consists of persons already familiar with basic nucleic acid hybridization assays, as well as those interested members of the first group. This group will benefit especially from the sections on applications. The chapters on applications address developments in veterinary and human medicine, as well as the often neglected markets of agriculture and environmental sciences, to shed light on the current and future direction of commercial and applied research interests.

This book was originally conceived as a primer on applications of nucleic acid technology, with an emphasis on the hybridization of nucleic acids, because of the central importance of this molecular interaction in defining the specificity and sensitivity of nucleic acid techniques in general. Obviously, other techniques besides hybridization protocols must be considered when developing applications, including those for commercial use. Various approaches to address sample handling and processing, user interface, and signal detection may be coupled to the basic hybridization reaction to design or to customize an efficient and potentially profitable application. The fashioning of a nucleic acid-based application is defined by utility and driven by ingenuity. Recognizing the broad spectrum of possible strategies, and the availability of several excellent techniques-oriented books (such as *Current Protocols in Molecular Biology*, or the abridged volume, *Short Protocols in Molecular Biology*, edited by F. M. Ausubel et al. and published by John Wiley & Sons), this text does not attempt to present detailed protocols for specific applications, but instead serves as a menu of options with examples describing representative applications. It is hoped that by illustrating different fields that might benefit from the development of assays, and by clarifying some of the basic technology, this text will

help stimulate the development of novel and commercially successful applications of nucleic acid hybridization techniques.

The contributions, cooperation, and patience of many authors were required in the preparation of this volume, and I am indebted to all of them. Dr. Bennie I. Osburn and Peter Prescott were integral in the initial conception and framing of this book. I am deeply grateful for their advice and assistance. Susan King and Colette Bean, both of John Wiley & Sons, shepherded this project to completion. I thank them for their endurance, guidance, and support.

CHARLES A. DANGLER

CONTENTS

Contributors vii

Preface
Charles A. Dangler ix

1. **Nucleic Acid Hybridizations: Principles and Strategies**
Asim K. Bej 1

2. **Generation of Nucleic Acid Probe Molecules**
Chitrita Debroy and Charles A. Dangler 31

3. **Nonisotopic DNA Labeling Strategies**
Calvin P. H. Vary 47

4. **Approaches to Genetic Analysis: Genetic Disease and Screening**
Linda L. McCabe and Edward R. B. McCabe 67

5. **Approaches to Genetic Analysis: Forensics**
Catherine Theisen Comey and Bruce Budowle 79

6. **Approaches to Infectious Disease Diagnosis**
William C. Wilson 105

7. **Diagnostic Nucleic Acid Probe Technology in Veterinary Medicine**
Jarasvech Chinsangaram and Bennie I. Osburn 131

8. **Diagnostic Nucleic Acid Probes in Human Medicine: Applications, Present and Future**
James Versalovic and James R. Lupski 157

9. **Developments in Food Technology: Applications and Regulatory and Economic Considerations**
Peter Feng, Keith A. Lampel, and Walter E. Hill 203

10. **Current Development and Applications of Nucleic Acid Technology in the Environmental Sciences**
Asim K. Bej and Meena H. Mahbubani 231

Index 275

CHAPTER 1

NUCLEIC ACID HYBRIDIZATIONS: PRINCIPLES AND STRATEGIES

ASIM K. BEJ
Department of Biology,
University of Alabama, at Birmingham,
Birmingham, AL 35294-1170

1.1. INTRODUCTION

The study of the myriad arrangements of the four nucleotides that constitute the genomes of each and every organism in this diversified living world has had a profound influence in virtually all areas of biology. The genetic make-up of each living organism contains a vast library of information. The analysis of these immense genetic messages requires the isolation and characterization of particular DNA sequences of interest. Nucleic acid hybridization is one of the primary approaches by which a great deal of information about the genetic structures and functions in all living organisms is being unfolded. The process that underlines all approaches involving nucleic acid hybridization, first described by Marmur and Doty [1], is the formation of the double helix from two strands of DNA that are complementary to each other. Successful nucleic acid hybridization is based on two fundamental criteria: (1) the two DNA strands involved in the hybridization process must have some degree of complementarity and (2) following hybridization, the extent of complementarity will determine the stability of the duplex DNA. Based on these two underlying principles of nucleic acid hybridization, a variety of avenues have been established involving this unprecedented approach to unveil the complexity of the underlying genetic principles of the living world and to address numerous enigmatic biological problems. This chapter will portray the fundamental principles of the

Nucleic Acid Analysis: Principles and Bioapplications, pages 1–29

nucleic acid hybridization method and highlight its role in some of the portentous areas of the biological sciences.

1.2. WHAT IS NUCLEIC ACID HYBRIDIZATION?

Nucleic acid hybridization is a process in which inconsonant nucleic acid strands with specific organization of nucleotide bases exhibiting complementary pairing with each other under specific given reaction conditions, thus forming a stable duplex molecule. This phenomenon is possible because of the biochemical property of base-pairing, which allows fragments of known sequences to find complementary matching sequences in an unknown DNA sample. The known sequence is called the probe, which is allowed to bind and hybridize with a complementary target sequence under specific conditions and provide a signal to confirm the presence of a specific sequence of interest in a sample. For hybridization, both the probe and the target DNA have to be in single-stranded form. Since the DNA molecule *in vivo* or *in vitro* exists in double-stranded form, the strands must be "denatured." By reversing the denaturation process, the probe is allowed to bind to its complementary target strand by a process called "reassociation" or "annealing." During duplex formation between the probe and the target strands, by the annealing process, a hybrid double-stranded molecule is formed, and this reaction is called "hybridization." Conceptually, all nucleic acid hybridization methods appear to be the same. However, actual performance and interpretation of each hybridization reaction involving binding of a unique probe with its complementary target DNA requires evaluation and optimization of a set of parameters to resolve the reaction conditions.

Since both DNA and RNA share the same biochemical principles of base-pairing, the hybrid formation during the nucleic acid hybridization process can occur between DNA–DNA, DNA–RNA, or RNA–RNA. Also, a number of approaches such as dot blot, colony or plaque lifts, Southern blot (DNA–DNA), Northern blot (RNA–DNA), *in situ,* and solution hybridization (all based on the basic principles of nucleic acid hybridization) are used.

1.3. SAMPLE PREPARATION FOR NUCLEIC ACID HYBRIDIZATION

The preparation of the target nucleic acids for the probe to bind is particularly important for *in vitro* hybrizations. The target DNA is intimately associated with complex proteins, lipids, carbohydrates, and other nucleic acids within the cell. Therefore, the release of the nucleic acids from the cells, with successive purification from complex cellular matrices, followed by denaturation, is the first and primary step for a successful hybridization reaction. In general, the cells are lysed by treating with detergent, alkali, or simply boiling [2,3]. In some hybridization approaches, such as dot blot, Southern blot, Northern blot, or solution hybridization, purification of the nucleic acids following lysis is recommended. However, in many

hybridization methods such as colony or plaque lifts, the cells are lysed, nucleic acids are denatured, and hybridization is performed without further purification. The *in situ* hybridization approach allows the probe to bind with previously denatured target nucleic acids without disrupting the tissues, cells, or chromosome [4].

1.4. VARIOUS APPROACHES AND KINETICS OF NUCLEIC ACID HYBRIDIZATIONS

There are three approaches to nucleic acid hybridization: hybridization in liquid phase (solution hybridization), hybridization on solid supports, and *in situ* hybridization. They are commonly used depending on the type of information that is expected to be generated from the hybridization result. These three hybridization approaches are described below.

1.4.1. Solution Hybridization

In the liquid hybridization, both the target nucleic acid and the probe interact in an optimized hybridization solution. The hybridization in the liquid phase has the maximum kinetics. Therefore, most of the time, the kinetics of a hybridization reaction is determined in liquid phase. Also, analysis of genome organizations and genetic complexity are generally done in liquid phase hybridizations. In liquid hybridization, the nucleic acids are purified from the target cells or tissues and transferred to a hybridization solution in which the single-stranded probe is added for hybridization. The nucleic acids are then denatured either by heating the sample or by treatment with alkali, and the annealing of the probes to their complementary target nucleic acids is performed. Hybridization may be followed by monitoring continuous decrease in the optical density, in a spectrophotometer at 260 nm wavelength, as the annealing of the probe to its complementary target nucleic acid occurs. The rate of the reaction is a measure of the concentration of the complementary sequences.

In liquid phase and possibly other hybridization processes, the initial event in the annealing of the two complementary strands is a nucleation reaction followed by a rapid process of "zippering" of the two single-stranded nucleic acids forming a duplex molecule. A mathematical model based on the assumption that the two strands are unbroken and have 100% complementarity was derived by Wetmur and Davidson [5]. However, in practice, unless the complexity of the genome is simple and the size is small such as in viruses, the complex nucleic acid molecules from eukaryotes or other higher organisms are sheared to a desired size range before being subjected to solution hybridization. As a result, during hybrid formation, numerous single-stranded tails are available for binding with other molecules, forming complex concatamers. The remaining single-stranded forms are one of the key features for quantitative determination of the hybrid molecules. Conventionally, following hybridization the sample is subjected to S1 nuclease digestion, which destroys the single-stranded molecules. The sample is then passed through a hydroxyapatite chromatography column in which the double-stranded duplex mole-

cules are selectively bound and the single-stranded molecules are eluted by using a buffer with specific ionic strength. Later, the double-stranded molecules are eluted using another buffer with a different ionic strength, and quantitation is determined by spectrophotometric, flurometric, colorimetric, or radioactive methods. An alternative approach following hybridization and S1 nuclease treatment is selective precipitation of the duplex molecules using trichloroacetic acid (TCA). Hydroxyapatite binding measures the fraction of the DNA that is linked to structures containing duplexes and follows the equation [6]

$$H = (1 + kC_0t)^{-1}$$

where H is fraction of DNA bound to the hydroxyapatite, k is observed rate constant ($M \cdot sec^{-1}$), C_0 is original concentration of the nucleotides ($mol \cdot L^{-1}$), and t is time (in seconds). In 0.14 M sodium phosphate buffer with a pH of 6.8, the double-stranded DNA will bind to hydroxyapatite, but the single-stranded DNA will not.

During the S1 nuclease digestion the single- and double-stranded nucleic acids will be distinguished, and the kinetics of the nucleic acid hybridization will follow the equation

$$S = (1 + kC_0t)^{-0.44}$$

where S is fraction of nucleotides that remain unpaired, and k, C_0, and t are the same as above.

If the DNA is sheared to a desired size range, the concentration of the fragments that contain a particular sequence is inversely proportional to the genome size. As a result, the rate of the reannealing reaction between these strands is inversely proportional to the genome size. This is more pronounced in eukaryotic organisms because of the presence of a large number of repetitive sequences in the genome [5,7]. The relationship between the complexity of the genome and the hybridization reaction rate in solution is fundamental to both DNA–DNA and DNA–RNA hybridizations. In the RNA–DNA hybridization, the base sequence complexity of a species of RNA can be defined as the base sequence complexity of the DNA from which the RNA is transcribed. Therefore, fundamentally the DNA–DNA hybridization can provide a comparative picture of the genome complexity of a group of related or unrelated organisms when performed in solution. On the other hand, a quantitative evaluation of the expression of a specific gene requires identification of the specific message by hybridization followed by measurement of the total amount of the duplex molecules.

In solution hybridization, the complexity of the genome can be measured quantitatively by calculating the C_0t (for DNA) or R_0t (for RNA) value. C_0t (moles of nucleotide per liter) is the product of DNA concentration (R_0t for the RNA concentration) and time (in seconds) of incubation in a reannealing reaction. The complexity is presented as a measure of $C_0t_{1/2}$, which is a value when one-half (50%) of the total DNA in the reaction has reannealed. The C_0t value can be determined from the equation

$$C/C_0 = 1/1 + kC_0t_{½}$$

where C is concentration of nucleotides at time t (mole per liter), C_0 is original concentration of nucleotides at time 0 (mole per liter), k is reassociation rate constant, and t is time (in seconds). Therefore, $C_0t_{½}$, i.e., when the reannealing reaction is half completed at $t_{½}$, can be defined mathematically as

$$C/C_0 = 1/2 = 1/1 + kC_0t_{½}$$

so that

$$C_0t_{½} = 1/k.$$

Since the $C_0t_{½}$ is the product of the concentration and the time required to proceed halfway, a greater $C_0t_{½}$ implies a slower reaction, which indicates the degree of complexity of the genome. Recently, high performance liquid chromatography (HPLC) has been used to separate and quantitate double-stranded hybrids from the single-stranded molecules following solution hybridization rapidly and accurately.

The complexity of any DNA can be determined by comparing its $C_0t_{½}$ with that of a standard DNA of known complexity. Usually *Escherichia coli* DNA is used as a standard. Its complexity is taken to be identical with the length of a genome that implies that every sequence in the *E. coli* genome of 4.2×10^6 is unique. Therefore, the complexity of any DNA can be determined by the equation

$C_0t_{½}$ (DNA of any genome)/$C_0t_{½}$ (*E. coli* DNA) =
Complexity of any genome/4.2×10^6 bp

Comprehensive descriptions of methods for C_0t [8,9] and R_0t [10] are given elsewhere.

The primary advantage of the solution hybridization approach is that it provides kinetic simplicity with maximum rate of hybridization. One of the disadvantages of this approach is that the nucleic acids should be free of proteins, lipids, carbohydrates, and other contaminants that may interfere with the annealing process. Also, because of its faster reassociation rate, higher self-annealing may occur. To prevent this problem, excess amount of probe may be used in a solution hybridization reaction. This approach may not be efficient if a large number of samples have to be analyzed.

1.4.2. Hybridizations on Solid Support

In this approach, the single-stranded DNA or RNA is immobilized on a solid support such as nitrocellulose, nylon, or polystyrene and is available for hybridization with the probe in a liquid phase. After it was first described by Gillespie and Speigelman [11], the hybridization method on solid support became a popular ana-

lytical hybridization tool in molecular biology despite the fact that hybridization in the two phases (solid and liquid, sometimes referred to as "mixed phase") can be slow. Since the target nucleic acids are bound on the solid support, one of the advantages of this approach is that self-annealing is prevented. The probe molecules are labeled radioactively or nonradioactively so that following the hybridization process the hybrids can be detected either by autoradiography or by colorimetry. Also, a quantitative assay of the hybrids can be determined by using a scintillation counter or a densitometer.

Types of Solid Supports. A variety of solid supports for DNA–DNA or RNA–DNA hybridizations are available. Among some of the membrane solid supports, nitrocellulose, nylon, or chemically activated papers are common. The nature and type of membrane solid supports are selected depending on the purpose of the hybridization.

Nitrocellulose membrane binds with DNA and RNA with high efficiencies (80 $\mu g/cm^2$). However, nucleic acid fragments below 500 nucleotides in length are bound poorly. Also, the nitrocellulose membrane is relatively fragile and becomes brittle. Therefore, it is difficult to handle, especially when used multiple times.

The introduction of nylon membrane for solid support hybridizations has overcome this difficulty. The texture of the nylon membranes is more flexible, and they do not disintegrate with repeated usage. Moreover, nucleic acids of sizes lower than 500 nucleotides can be immobilized efficiently on the nylon membranes. The sensitivity of hybridization detection on various commercially available nylon membranes is claimed to be much higher than on the nitrocellulose membranes. For Southern and Northern hybridizations, both nylon and nitrocellulose membranes provide excellent hybridizations. For colony lifts, a nylon membrane with a pore size of 1.2 μm is recommended. For DNA dot blots, a membrane with a 0.45 μm pore size is recommended for large (>500 nucleotide) molecules, and a 0.22 μm pore size membrane is preferred for small (<500 nucleotide) molecules. For RNA dot blots, membranes with pore sizes of 0.1–0.22 μm provide excellent hybridization.

Recently, the use of a polystyrene surface (e.g., microtiter plates) has been used for hybridizations. Although this approach has not become popular to date, some of the advantages are (1) multiple samples can be handled more easily on a microtiter plate than on a nylon or nitrocellulose membrane, (2) automation is possible in this approach, (3) less amounts of reagents are used in a microtiter plate hybridization, and (4) the hybridization time can be shortened by several hours compared with membrane hybridization. The efficiency of hybridization in a microtiter plate has not been shown to be higher than in the membrane hybridization.

Immobilization of Nucleic Acids on Solid Supports. As a first step the nucleic acids are denatured before immobilizing onto solid supports. The conventional methods of immobilization do not allow the nucleic acids to bind to the nitrocellulose or nylon membranes covalently. The noncovalent immobilization of the nucleic acids may pose some problems such as at high temperature the nucleic acids may

leach out from the membrane if hybridization is carried on for a long period of time. In addition, if the probe binds to the immobilized target nucleic acid by its entire length, the hybrid molecule may detach from the membrane and stay in the liquid phase [12]. This phenomenon may lead to no or unexpectedly poor hybridization signals with time. Therefore, it is necessary to bind the immobilized nucleic acids onto the membrane covalently. Most of the commercially available membranes are designed to bind to the nucleic acids covalently either by UV irradiation or by chemical treatment [13]. The chemically activated membranes bind with the nucleic acids covalently. Therefore, these membranes can be used for immobilization of both large and small nucleic acids. However, the binding capacity of the chemically activated membranes is much lower (1–2 $\mu g/cm^2$) than that of nitrocellulose membranes and many of the commercially available nylon membranes, and the procedures for binding are more complicated. The chemically activated membranes are commonly used for hybrid selection for the enrichment of specific RNA sequences [10].

Various Types of Solid Support Hybridizations. The solid support hybridization approach has broad applications, including (1) dot blot or slot blot hybridization, (2) bacterial colony or plaque hybridization, (3) Southern hybridization, (4) Northern hybridization, (5) *in situ* hybridization, and (6) hybrid selection by sandwich hybridization.

Dot Blot or Slot Blot Hybridization. The dot or slot blot hybridization approach is relatively rapid with the advantage that a large number of samples can be analyzed simultaneously against one probe under the same hybridization conditions. The method involves fixation of the single-stranded target nucleic acid molecules onto a solid support, which is a nitrocellulose or a nylon membrane containing free primary amine groups available for covalent binding with the nucleic acid molecules. The nucleic acid samples are immobilized with specific geometric arrays manually or by using commercially available dot or slot blot apparatuses that are capable of analyzing 96 or more nucleic acid samples simultaneously on a single solid support. The name of this hybridization approach is derived from the geometric arrays of the samples on the solid support, which can be in the form of solid small circles (dot blot) or solid moderately elongated wells (slot blot). The rest of the solid support that does not contain any target nucleic acid to be tested is inactivated by treatment with a blocking reagent followed by hybridization with a specific nucleic acid probe under specific hybridization conditions. Following hybridization, the solid support is processed by washing several times with several solutions to remove unbound probes and/or binding of the probe with nontarget nucleic acids.

Although purified nucleic acids were initially used for dot or slot blot hybridization assays, a number of studies showed that unpurified nucleic acid from the environmental or clinical samples can be used successfully for hybridization simply by lysing the cells in the samples with a strong base or a chaotropic agent such as sodium iodide, denaturing the nucleic acid with alkali such as sodium hydroxide, and denaturing the cellular proteins by protein-degrading enzymes such as proteinase K