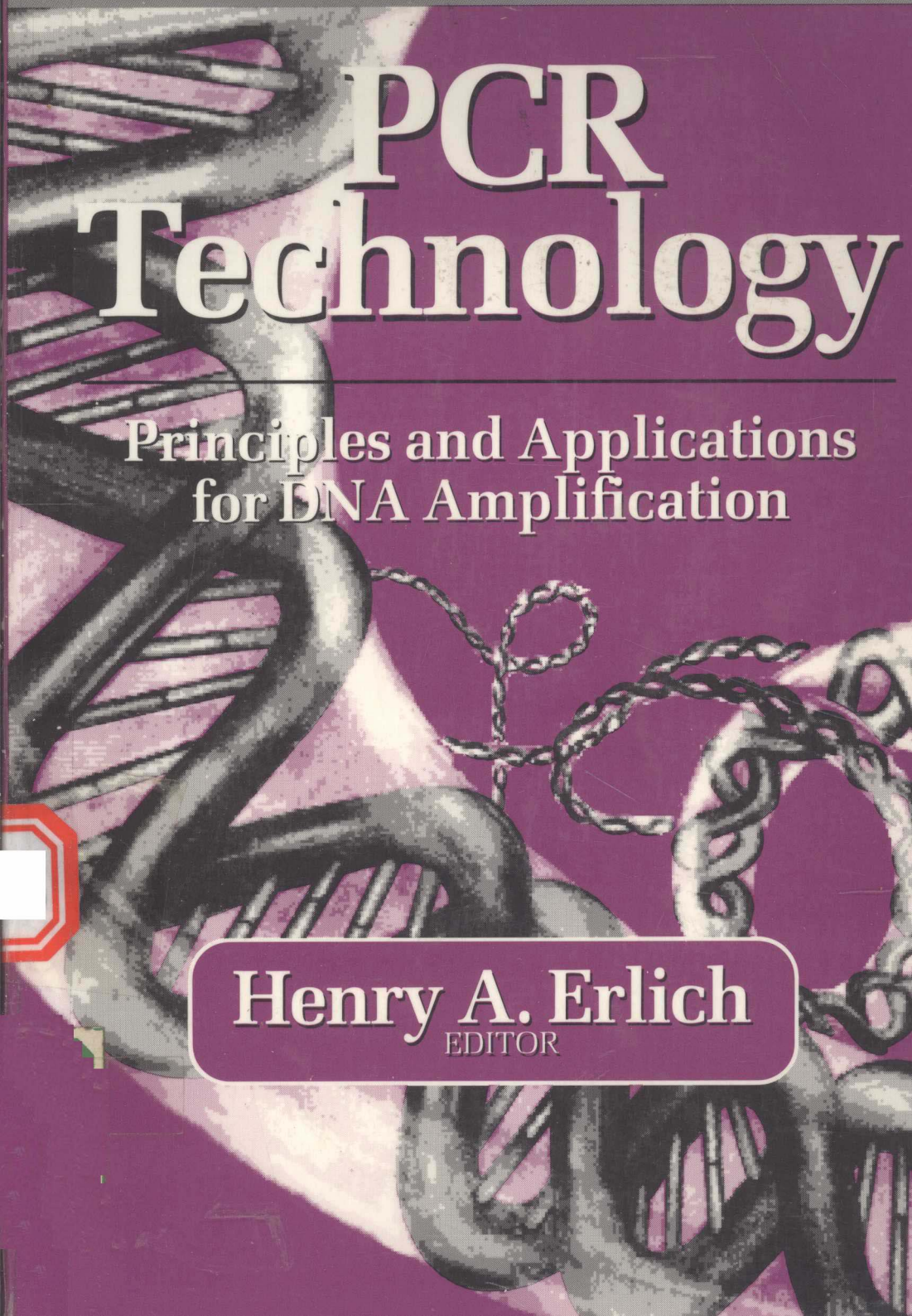
PCR Technology

Principles and Applications for DNA Amplification



Henry A. Erlich
EDITOR

PCR Technology

Principles and Applications for DNA Amplification

Henry A. Erlich
EDITOR

M
stockton press

New York London Tokyo Melbourne Hong Kong

For Brenda and Justin

Published in the United States and Canada by Stockton Press, 1989
15 East 26th Street, New York, N.Y. 10010

Library of Congress Cataloging-in-Publication Data

PCR technology.

(Breakthroughs in molecular biology)
1. DNA polymerases. 2. DNA—Synthesis.
3. DNA polymerases—Biotechnology. I. Erlich, Henry A., 1943- . II. Series. [DNLM: 1. DNA Polymerases. 2. Gene Amplification. 3. Genetic Engineering. 4. Molecular Biology. 5. RNA Polymerases. QH 442 P348]
QP606.D64P37 1989 574.87'3282 89-4342
ISBN 0-935859-56-X (pbk.)

Published in the United Kingdom by
MACMILLAN PUBLISHERS LTD (Journals Division), 1989
Distributed by Globe Book Services Ltd,
Brunel Road, Houndmills, Basingstoke,
Hants RG21 2XS, England

British Library Cataloguing in Publication Data

Erlich, Henry
DNA amplification using the PCR
1. Organisms. DNA polymerases
I. Title
574.87'3282

ISBN 0-333-48948-9

Printed in the United States of America
9 8 7 6 5 4 3 2

PREFACE

Since the first report of specific DNA amplification using the polymerase chain reaction (PCR) in 1985, the number of different applications has grown steadily, as have the modifications of the basic method. This increase has been claimed to be exponential - like the PCR itself. In any case, detailed discussion of the full range of biological problems under investigation as well as all of the various experimental approaches utilizing PCR is beyond the scope of this book. Conceived as a snapshot of a rapidly moving field, the book's contributed chapters have been organized into three sections: Basic Methodology, Research Applications, and Medical Applications. I have used the brief introduction to each section to cite some of the more recent literature relevant to each area and to "fill gaps" by directing the reader to topics not covered. Some chapters provide detailed protocols, listing favorite "PCR recipes," while others give an overview of a particular field. My hope is that an interested reader, armed with the knowledge of some of the methodological issues and of some of the applications, could proceed to devise PCR-based approaches relevant to his or her area of investigation.

I am grateful to my current and former colleagues at Cetus for their contributions to this book and, of course, above all, to the technology that occasioned it. To Kary Mullis, who invented the basic PCR method, and to my colleagues in the Human Genetics Department (Randall Saiki, Stephen Scharf, Glenn Horn, Fred Faloona, Kary Mullis, and Norm Arnheim) who first developed it for the analysis of human genetic variation. To David Gelfand and Susanne Stoffel who isolated and characterized the thermostable DNA polymerase (*Taq* polymerase) whose use in the PCR led to the current rapid and automated procedure and transformed the reaction from a method of last resort to one of first choice. To Corey Levenson, Dragan Spasic, and Lauri Goda who responded gracefully to the increased demand for synthetic oligonucleotides generated by the PCR. Also to Tom White who,

as Director of Research, saw PCR's potential and supported its development and to John Sninsky who, in addition to his contributions applying PCR to virus detection, was instrumental in the program to develop the *Taq* polymerase and automated thermocylers. Finally, this book would not have been possible without the patience and dedication of Kathy Levenson who worked closely with the authors and, with the help of Dean Grantham, transformed the sporadic trickle of manuscripts into camera-ready copy.

CONTRIBUTORS

James W. Ajioka, Ph.D.
Post-doctoral Fellow
Department of Genetics
Washington University School of Medicine

Norman Arnheim, Ph.D.
Professor of Molecular Biology
Department of Biological Sciences
University of Southern California

Edward T. Blake, D. Crim.
Forensic Science Associates

Johannes L. Bos, Ph.D.
Professor
Sylvius Laboratoria
University of Leiden, The Netherlands

Teodorica L. Bugawan
Research Associate
Department of Human Genetics
Cetus Corporation

C. Thomas Caskey, Ph.D.
Professor and Director, Institute for Molecular Genetics
Investigator, Howard Hughes Institute
Baylor College of Medicine

J.S. Chamberlain, Ph.D.
Post-doctoral Fellow
Institute for Molecular Genetics
Baylor College of Medicine

David R. Cox, M.D., Ph.D.
Associate Professor of Psychiatry and Biochemistry/Biophysics
University of California, San Francisco

B.L. Daugherty, M.S.
Research Biochemist
Department of Cellular and Molecular Biology
Merck, Sharp and Dohme Research Laboratories

Henry A. Erlich, Ph.D.
Senior Scientist, Director
Department of Human Genetics
Cetus Corporation

Dan Garza, Ph.D.
Post-doctoral Fellow
Department of Genetics
Washington University School of Medicine

David Gelfand, Ph.D.
Vice-President, Scientific Affairs
Product Development, PCR Division
Cetus Corporation

Richard A. Gibbs, Ph.D.
Post-doctoral Fellow
Institute for Molecular Genetics
Baylor College of Medicine

Ulf Gyllensten, Ph.D.
Department of Medical Genetics
Biomedical Center, Uppsala, Sweden

J. Han, Ph.D.
Research Fellow
Department of Cellular and Molecular Biology
Merck, Sharp and Dohme Research Laboratories

Daniel L. Hartl, Ph.D.
McDonnell Professor of Genetics
Department of Genetics
Washington University School of Medicine

Russell Higuchi, Ph.D.
Scientist
Department of Human Genetics
Cetus Corporation

Ernest Kawasaki, Ph.D.
Scientist
Department of Human Genetics
Cetus Corporation

CONTRIBUTORS

James W. Ajioka, Ph.D.
Post-doctoral Fellow
Department of Genetics
Washington University School of Medicine

Norman Arnheim, Ph.D.
Professor of Molecular Biology
Department of Biological Sciences
University of Southern California

Edward T. Blake, D. Crim.
Forensic Science Associates

Johannes L. Bos, Ph.D.
Professor
Sylvius Laboratoria
University of Leiden, The Netherlands

Teodorica L. Bugawan
Research Associate
Department of Human Genetics
Cetus Corporation

C. Thomas Caskey, Ph.D.
Professor and Director, Institute for Molecular Genetics
Investigator, Howard Hughes Institute
Baylor College of Medicine

J.S. Chamberlain, Ph.D.
Post-doctoral Fellow
Institute for Molecular Genetics
Baylor College of Medicine

Haig H. Kazazian, Jr., M.D.
Professor of Pediatrics, Gynecology, and Obstetrics
Genetics Unit, Department of Pediatrics
The Johns Hopkins Hospital

Thomas D. Kocher, Ph.D.
Post-doctoral Fellow
Department of Biochemistry
University of California, Berkeley

Shirley Kwok
Associate Scientist
Department of Infectious Diseases
Cetus Corporation

Simon W. Law, Ph.D.
Senior Research Fellow
Department of Cellular and Molecular Biology
Merck, Sharp and Dohme Research Laboratories

G.E. Mark, Ph.D.
Associate Director
Department of Cellular and Molecular Biology
Merck, Sharp and Dohme Research Laboratories

Richard M. Myers, Ph.D.
Assistant Professor
Department of Physiology and Biochemistry/Biophysics
University of California, San Francisco

David L. Nelson, Ph.D.
Post-doctoral Fellow
Howard Hughest Institute
Baylor College of Medicine

Howard Ochman, Ph.D.
Research Assistant Professor
Department of Genetics
Washington University School of Medicine

L. O'Neill, M.S.
Staff Biochemist
Department of Cellular and Molecular Biology
Merck, Sharp and Dohme Research Laboratories

Christian Oste, Ph.D.
Bioscience Specialist
Perkin-Elmer Cetus

Randall Saiki
Associate Scientist
Department of Human Genetics
Cetus Corporation

George Sensabaugh, Ph.D.
Professor and Chairman
Department of Public Health
University of California, Berkeley

Val C. Sheffield, M.D., Ph.D.
Post-doctoral Fellow
Department of Physiology and Biochemistry/Biophysics
University of California, San Francisco

John J. Sninsky, Ph.D.
Director
Department of Infectious Diseases
Cetus Corporation

J.-S. Tung, Ph.D.
Senior Research Fellow
Department of Cellular and Molecular Biology
Merck, Sharp and Dohme Research Laboratories

Cecilia H. von Beroldingen, Ph.D.
Post-doctoral Fellow
Department of Public Health
University of California, Berkeley

Alice Wang
Associate Scientist
Department of Molecular Biology
Cetus Corporation

Thomas J. White, Ph.D.
Senior Director of Diagnostics Research
Hoffmann La Roche Inc.

CONTENTS

PART ONE

BASIC METHODOLOGY

In the short history of molecular biology, the emergence of a new technique (e.g., Southern blotting, molecular cloning, pulsed field gel electrophoresis) has often transformed the way we think about approaching both fundamental and applied biological problems. The capacity to amplify specific segments of DNA, made possible by the polymerase chain reaction (PCR), represents such a change. The PCR is an *in vitro* method for the enzymatic synthesis of specific DNA sequences, using two oligonucleotide primers that hybridize to opposite strands and flank the region of interest in the target DNA. A repetitive series of cycles involving template denaturation, primer annealing, and the extension of the annealed primers by DNA polymerase results in the exponential accumulation of a specific fragment whose termini are defined by the 5' ends of the primers. Because the primer extension products synthesized in one cycle can serve as a template in the next, the number of target DNA copies approximately doubles at every cycle. Thus, 20 cycles of PCR yields about a million-fold (2^{20}) amplification. This method, which was invented by Kary Mullis,[1,2] was originally applied by a group in the Human Genetics Department at Cetus to the amplification of

human β-globin DNA and to the prenatal diagnosis of sickle-cell anemia.[3,4,5]

Initially, the PCR used the Klenow fragment of *E. coli* DNA polymerase I to extend the annealed primers. This enzyme was inactivated by the high temperature required to separate the two DNA strands at the outset of each PCR cycle. Consequently, fresh enzyme had to be added during every cycle. The introduction of the thermostable DNA polymerase (*Taq* polymerase) isolated from *Thermus aquaticus* (see Chapter 2) transformed the PCR into a simple and robust reaction which could now be automated by a thermal cycling device (see Chapter 3). The reaction components (template, primers, *Taq* polymerase, dNTP's, and buffer) could all be assembled and the amplification reaction carried out by simply cycling the temperature within the reaction tube.[6] The effect of varying the reaction parameters (e.g., enzyme, primer and Mg^{++} concentration as well as the temperature cycling profile) on the specificity and yield of the amplification is discussed in Chapter 1. Although, for any given pair of oligonucleotide primers, an optimal set of conditions can be established, there is no single set of conditions that will be optimal for all possible reactions.

The specificity of the PCR is typically analyzed by evaluating the production of the target fragment relative to other products by gel electrophoresis. Another factor influencing the homogeneity of the PCR product is the *concentration* of the target sequence in the genomic template.[6] This effect was revealed by experiments[6] in which normal genomic DNA (with two copies of the β-globin gene/cell) and genomic DNA with a homozygous deletion of β-globin were amplified with β-globin primers. Using the normal genomic template, this reaction generated a unique β-globin fragment with no detectable nonspecific products. Using the β-globin deletion DNA sample, no β-globin fragment was synthesized, as expected; however, several non-target fragments were produced. Thus, in the absence of the "right" template, the "wrong" sequences became amplified by the β-globin primers, illustrating the old adage, "Idle hands are the Devil's playthings." This effect accounts, in part, for the generally heterogenous gel profile of PCR products synthesized from a very rare template sequence, like HIV sequences which are present in only a small fraction of the cells sampled. As discussed in Chapter 1, modification of various PCR parameters to optimize the specificity of amplification may yield more homogenous products even in rare template reactions.

The initial PCR method based on DNA synthesis by the Klenow enzyme at 37°C was not highly specific. Although a specific target fragment could be amplified up to a million-fold, most of what was synthesized in the PCR was not, in fact, this target fragment. By cloning a β-globin amplification reaction and screening the individual clones with a β-globin probe to detect the target sequence and with one of primers to detect *any* amplified sequence, the specificity of the PCR was estimated to be ~1%.[7] Other primer pairs were somewhat more or less specific but, in general, the Klenow enzyme PCR was not a very specific reaction and required subsequent analysis with a specific hybridization probe[3,4] or, in some cases, with internal "nested primers"[1] to detect and characterize the amplified target sequence.

The use of the *Taq* polymerase not only simplified the PCR procedure (see above) but significantly increased the specificity and the overall yield of the reaction. The higher temperature optimum for the *Taq* polymerase (~75°C) allowed the use of higher temperatures for primer annealing and extension, thereby increasing the overall stringency of the reaction and minimizing the extension of primers that were mismatched with the template. At 37°C, many of these mismatched primers are sufficiently stable to be extended by the Klenow enzyme, resulting in non-specific amplification products. The increase in the specificity of the *Taq* PCR results in an improved yield of the amplified target fragment by reducing the competition by non-target products for enzyme and primers. In the later cycles, the amount of enzyme is no longer sufficient to extend all the annealed primer/template complexes in a single cycle period, resulting in a reduced efficiency and a "plateau" in the amplification reaction. This plateau is reached later (e.g., about 30 cycles rather than 20 starting with 1 µg of genomic DNA) in the *Taq* PCR than in the reaction with the Klenow enzyme due to the increased specificity of the former reactions. Other factors like the reassociation of the template strands at high product concentration may also contribute to the plateau effect and are discussed in Chapter 1. In addition to the increase in the specificity and yield of PCR made possible by *Taq* polymerase, the use of this enzyme allows the amplification of much longer fragments (up to 10 kb, albeit with reduced efficiency)[8] than does the Klenow enzyme (<400 bp).

Although the PCR is considered primarily a method for producing copies of a specific sequence, it is also a very powerful and precise way of altering a particular template sequence. Since the oligonucleotide primers become physically incorporated into the amplified product and mismatches between the 5' end of the primer and *initial* template[†] are tolerated, it is possible to introduce new sequence information adjacent to the target sequence via the primers. Thus, for cloning a given sequence, one is no longer constrained by the restriction sites Nature provides but can add any restriction enzyme recognition sequence to the 5' ends of the primer[7] leading to the formation of a new restriction site in the double-stranded amplification product. Similarly, regulatory elements such as a T7 promoter can be added,[1] allowing the synthesis of RNA copies from the PCR product using T7 RNA polymerase. Furthermore, specific nucleotide substitutions, insertions, and deletions can also be introduced into the amplified product with the appropriate primers.

Unlike these directed mutations, changes introduced into the sequence of the PCR products due to nucleotide misincorporation can create potential problems. An estimate of the fidelity of *Taq* polymerase during the PCR was made by the sequence analysis of multiple M13 clones generated by cloning the products of a particular amplification reaction. In this initial study,[6] the *frequency* of errors was found to be ~1/400 and the error *rate*

[†]After the first few cycles, virtually all of the templates have been synthesized in previous cycles and, therefore, contain the primer sequences.

was calculated to be ~$2x10^{-4}$ nt/cycle. More recent studies, carried out on a different gene using the current, optimized protocol (reduced Mg^{++} and dNTP concentrations), found a lower frequency of errors.[9] The original estimate is in reasonable agreement with the report of a 10^{-4} nt error rate determined by the *in vitro* replication of a β-galactosidase template by the *Taq* polymerase.[10] Although different template sequences may have a somewhat different "mutability" and different reaction conditions may influence the fidelity of the *Taq* polymerase, the original "high" error rate estimated for *Taq* polymerase PCR (~10^{-4}) does not pose a problem for most applications. In the analysis of the *population* of amplified products, as in oligonucleotide probe hybridization or in direct sequencing, the rare errors in individual products are not detectable. However, in the sequence analysis of individual clones derived from a PCR, sequences must be determined from multiple clones to distinguish misincorporated nucleotides from the faithful copes of the template sequence. Other DNA polymerases with an active "proof-reading" function, like T4 DNA polymerase, can carry out PCR[11] and may prove useful in studies where a very low error rate is required. As discussed in Chapter 2, the *Taq* polymerase does not contain measurable 3'-to-5' exonuclease "proof-reading" activity.

An important property of the PCR, particularly in diagnostic applications, is the capacity to amplify a target sequence from crude DNA preparations[4] as well as from degraded DNA templates.[12] The DNA in a sample need not be chemically pure to serve as a template provided that the sample does not contain inhibitors of *Taq* polymerase. The effect of various rapid protocols for sample preparation on the PCR is discussed in Chapter 4. The ability to amplify specific sequences from crude DNA samples has important implications for research applications, (e.g., sperm lysates[13]; see Chapter 12), for medical diagnostic applications (e.g., mouthwash[14] or archival paraffin-embedded tissue samples[15]) and for forensics (e.g., individual hairs[16]; see Chapter 17).

The possibility of contamination in the amplification reaction is an issue with broad implications for both research and diagnostic applications. Given the capacity of PCR to synthesize millions of DNA copies, contamination of the sample reaction with either products of a previous reaction (product carryover) or with material from an exogenous source is a potential problem - particularly in those reactions initiated with only a few templates. The amplification of individual sperm (see Chapter 12), of single hairs (see Chapter 17), and of HIV genomic sequences (see Chapter 19) all require rigorous measures to minimize and monitor potential contamination. In general, careful laboratory procedure, prealiquoting reagents, the use of positive displacement pipettes, and the physical separation of the reaction preparation from the analysis of the reaction products are all precautions that reduce the risk. Carrying out only the minimal number of PCR cycles required for analysis also minimizes the chance that a rare contaminating template will be amplified. A panel of "blank" reactions with no template DNA is necessary to detect potential contamination. In genetic typing, a sample that has been contaminated can often be identified by a genotyping result with more than 2 alleles (see Chapters 16 and 17).

The PCR, like recombinant DNA technology, has had an enormous impact in both basic and diagnostic aspects of molecular biology because it can produce large amounts of a specific DNA fragment from small amounts of a complex template. Recombinant DNA techniques create molecular clones by conferring on a specific sequence the ability to replicate by inserting it into a vector and introducing the vector into a host cell. PCR represents a form of "*in vitro* cloning" that can generate, as well as modify, DNA fragments of defined length and sequence in a simple automated reaction.

In its early, heady days, Erwin Chargaff once described molecular biology dismissively as "the practice of biochemistry without a license." Its spectacular success over the last 30 years suggests that significant progress can be made even in the absence of an established theoretical foundation. One element in the rapid development of molecular biology has been its characteristic embrace of new techniques and the aggressive pursuit of the new questions made possible by their use. PCR has facilitated molecular analysis in many disparate fields of biological research; given the ease and simplicity of PCR amplification, it can be said to allow the practice of molecular biology without a permit. It is likely that future practitioners will continue to find new applications and modifications of this powerful technology.

REFERENCES

1. Mullis, K.B., and Faloona, F. (1987) *Meth. Enzymol.* 155:335.
2. Mullis, K.B., Faloona, F., Scharf, S.J., Saiki, R.K., Horn, G.T., and Erlich, H.A. (1986) *Cold Spring Harbor Symp. Quant. Biol.* 51:263-273.
3. Saiki, R., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich, H.A., and Arnheim, N. (1985) *Science* 230:1350.
4. Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1986) *Nature* 324:163.
5. Embury, S.H., Scharf, S.J., Saiki, R.K., Gholson, M.A., Golbus, M., Arnheim, N., and Erlich, H.A. (1987) *New Engl. J. Med.* 316:656.
6. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S., Higuchi, R.H., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988) *Science* 239:487.
7. Scharf, S.J., Horn, G.T., and Erlich, H.A. (1986) *Science* 223:1076.
8. Jeffreys, A.J., Wilson, V., Neumann, R., and Keyte, J. (1988) *Nucl. Acids Res.* 16:10953-10971.
9. Goodenow, M., Huet, T., Saurin, W., Kwok, S., Sninsky, J., and Wain-Hobson, S. (1989) *J. Acquired Immune Deficiency Syndromes*, in press.
10. Tindall, K.R., and Kunkel, T.A. (1988) *Biochem.* 27:6008-6013.
11. Keoharong, P., Kat, A., Cariello, N.F., and Thilly, W.G. (1988) *DNA* 7:63-70.
12. Bugawan, T.L., Saiki, R.K., Levenson, C.H., Watson, R.M., and Erlich, H.A. (1988) *Bio/Technology* 6:943.
13. Li, H., Gyllensten, U.B., Cui, X., Saiki, R.K., Erlich, H.A., and Arnheim, N. (1988) *Nature* 335:414.
14. Lench, N., Stanier, P., and Williamson, R. (1987) *Lancet* ii:1356-1358.
15. Shibata, D.K., Martin, J.W., and Arnheim, N. (1988) *Cancer Res.* 48:4564-4566.
16. Higuchi, R., von Beroldingen, C.H., Sensabaugh, G.F., and Erlich, H.A. (1988) *Nature* 332:543