PCR Technology: An Introduction to Current Trends and Innovations

The first edition of *PCR Technology* was published almost 10 years ago. It is now possible to achieve throughput that is three orders of magnitude higher, amplify fragments that are two orders of magnitude longer, and quantify target copy numbers with an accuracy that is at least one order of magnitude higher than at the time when the previous edition was written. In addition, a large number of commercial kits and additives are now available, promising (and, for the less experienced researcher, often delivering) superior performance compared to “homemade” assays. However, many basic parameters of the reaction remain poorly understood. These include the kinetics and thermodynamics of the reaction, and the mechanisms by which some substances inhibit and others facilitate amplification. Optimization of reaction components and cycling steps is critical for the success of any new protocol, but the procedures used have hardly changed since PCR was invented. Commercial kits can save valuable time, but convenience and lack of sufficient information about the composition of these reagents might stifle further development.

This book contains detailed PCR protocols which readers of diverse background and expertise will find useful. These days, no single volume can provide an exhaustive collection of “recipes” covering all applications and variants of the technique. Rather, I have tried to provide examples of the current range of PCR strategies and uses in a way that I hope will encourage those familiar with the basic principles and practice to develop or improve their own protocols.

PCR has, no doubt, established itself as one of the most commonly used techniques in the life sciences. But the numbers of papers per year mentioning PCR now remain at a constant level or are even slightly decreasing (Figure 1). It is interesting to compare this trend with that in a somewhat older key technology of the life sciences, monoclonal antibodies. For both, there was a rapid increase in the number of published papers in the first decade after their introduction (1974 to 1986 for monoclonals, 1985 to 1995 for PCR). Thereafter, the output slowed down — most notably when compared to the growing total number of papers published in their respective areas (Figure 1 and Figure 2). At least for PCR, all major improvements in the technology so far were made during the first 10 years (Figure 2). Searches for papers containing the respective keywords revealed a similar decrease in the fraction of papers as for PCR in general (Figure 1B and not shown). An interesting exception is the “real-time PCR,” a relatively young invention.

Still, PCR remains the most popular technology in the life sciences, and for the moment there is no serious contender in sight. Isothermal amplification methods can be performed on simpler instruments. However, this advantage is offset by their need for more complex reaction mixtures, resulting in higher reagent costs and decreased robustness. This makes these methods less attractive for high throughput and quantitative applications. Quantum-dot probes now allow the detection of single molecules and polymorphisms without prior amplification, however, the method is limited to analytical applications. Thus, the unique strength of PCR remains its versatility, e.g. the same instruments and related expertise can be used for a large variety of demands. These will remain high for some time since the genome sequences of several key organisms have just recently been completed. Polymorphism and expression data obtained with the aid of these sequences are expected to yield a wealth of clinical and commercially important data, and PCR is at the forefront of the “genomics revolution.” It will be interesting to see whether this will lead not only to further specialization, but also major developments in the technique.
The book has three main sections. The first deals with the preparation for a PCR experiment: sample collection and pretreatment, optimization of the reaction, and primers for efficient labeling and cloning. Although not strictly necessary in every situation, these are general considerations applicable to most PCR experiments. The second section contains examples of "analytical" PCR, divided into qualitative and quantitative applications (e.g., those addressing the questions "Is a certain target sequence present or not?" and "How much of a target sequence does my sample contain?"). Because the recent years have seen major advances in high throughput, open systems, and quantitative PCR, each of these subsections is introduced by a review-style "overview" chapter.

The third section presents examples of "preparative" methods addressing the general problem of how to generate samples for further analysis, and in vitro evolution.

**FIGURE 1** Number of journal articles published per year containing the phrase "polymerase chain" compared to the words "monoclonal" and "antibody." The data shown were obtained with the search engine Scirus (www.scirus.com) which searches complete HTML and PDF documents on the web. I found similar results with PubMed (http://www.ncbi.nlm.nih.gov/entrez/), a database that allows searching in title, keywords, or abstract only. Both sources suggest that the published output reached saturation for both technologies about 10 years after their introduction (A). Owing to a general and continuous increase in the number of life science papers, this trend is more obvious for the relative numbers (B).
Inadvertently, some chapters fit more than one category. The chapter by Maito Remm et al. on arrayed primer extension PCR contains useful information on primer design that is relevant to all PCR methods and could have otherwise contributed to a separate chapter in the first section. Likewise, the chapter by Xiao Yan Zhong et al. contains a protocol for the Taqman assay that complements the section on quantitative PCR, as well as the chapter on whole genome amplification by Dagan Wells and Mercedes G. Bermudez in the “preparative” section.

Like many great inventions, the polymerase chain reaction is deceptively simple in principle, amazingly versatile in its uses, but frustratingly difficult to master in its more advanced forms. Accordingly, the technical difficulty of the protocols in this book varies. Table 1 contains a breakdown of those characteristics that can make a PCR experiment either “easy-PC” or arcane “R-T.” Novices should avoid as many of the “difficult” elements as possible or it might appear that this apparently so robust technique is under the spell of voodoo or the moon phases! However, useful tips for the beginner can be hidden even in the more advanced chapters.

Our theoretical understanding of the PCR process is still in its infancy. How much of the behavior is controlled by equilibrium thermodynamics, how much by the kinetics of the reaction? How do the various components of the reaction mixture interact to determine template DNA melting, primer annealing, and polymerase activity? Nevertheless, a number of attempts have been made to model aspects of the reaction, some of which are included in this book (Table 2).

I hope that this small volume will stimulate readers of all levels to “pick and mix” from different protocols and develop their own innovative approaches that will take PCR into a successful third decade.

FIGURE 2  Percentage of journal papers mentioning “polymerase chain” sorted by scientific discipline (Scirus, www.scirus.com). Major innovations in the technique are shown above the figure.

Inadvertently, some chapters fit more than one category. The chapter by Maito Remm et al. on arrayed primer extension PCR contains useful information on primer design that is relevant to all PCR methods and could have otherwise contributed to a separate chapter in the first section. Likewise, the chapter by Xiao Yan Zhong et al. contains a protocol for the Taqman assay that complements the section on quantitative PCR, as well as the chapter on whole genome amplification by Dagan Wells and Mercedes G. Bermudez in the “preparative” section.

Like many great inventions, the polymerase chain reaction is deceptively simple in principle, amazingly versatile in its uses, but frustratingly difficult to master in its more advanced forms. Accordingly, the technical difficulty of the protocols in this book varies. Table 1 contains a breakdown of those characteristics that can make a PCR experiment either “easy-PC” or arcane “R-T.” Novices should avoid as many of the “difficult” elements as possible or it might appear that this apparently so robust technique is under the spell of voodoo or the moon phases! However, useful tips for the beginner can be hidden even in the more advanced chapters.

Our theoretical understanding of the PCR process is still in its infancy. How much of the behavior is controlled by equilibrium thermodynamics, how much by the kinetics of the reaction? How do the various components of the reaction mixture interact to determine template DNA melting, primer annealing, and polymerase activity? Nevertheless, a number of attempts have been made to model aspects of the reaction, some of which are included in this book (Table 2).

I hope that this small volume will stimulate readers of all levels to “pick and mix” from different protocols and develop their own innovative approaches that will take PCR into a successful third decade.
## TABLE 1
Easy and Difficult Elements in PCR Protocols

<table>
<thead>
<tr>
<th>Easy</th>
<th>Difficult</th>
<th>Relevant Chapters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>Multiplex (&gt;2)</td>
<td>Not discussed in this book</td>
</tr>
<tr>
<td>Difficult</td>
<td>Whole genome amplification</td>
<td>Wells, Zhong et al.</td>
</tr>
<tr>
<td>Qualitative</td>
<td>Accurate qualitative</td>
<td>Bustin, Vu, Dear</td>
</tr>
<tr>
<td>Specific primers</td>
<td>Degenerate primer(s)</td>
<td>Not discussed in this book</td>
</tr>
<tr>
<td>Low throughput</td>
<td>High throughput</td>
<td>Mir and Ragoussis, Butler and Vallone, Tost et al., Remm et al.</td>
</tr>
<tr>
<td>In tube or well (20–100 µl)</td>
<td>In situ</td>
<td>Nuovo, Mennick et al., Mage</td>
</tr>
<tr>
<td>Small amplicons (100–600 bp)</td>
<td>Long PCR (&gt;2kb)</td>
<td>Fromenty et al.</td>
</tr>
<tr>
<td>Known sequence</td>
<td>(Partially) unknown sequence</td>
<td>Hui and Lo, Myrick, Garcés et al., Allander et al.</td>
</tr>
<tr>
<td>Target sequence</td>
<td>Stable secondary structure (“GC-rich”)</td>
<td>Chakrabarti, Weissensteiner</td>
</tr>
<tr>
<td>Easily denatured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intact nucleic acids</td>
<td>Degraded DNA</td>
<td>Stone, Fromenty et al.</td>
</tr>
<tr>
<td>High target copy numbers (&gt; 10,000)</td>
<td>Low target copy numbers (1–10)</td>
<td>Dear, Wells, Zhong et al.</td>
</tr>
<tr>
<td>High ratio of target to background sequence (cDNA, simple genomes)</td>
<td>Presence of large amounts of nontarget sequence (complex genomes, somatically mutated cells, viral load)</td>
<td>Radstrom et al., Allander et al., Nuovo, Mage</td>
</tr>
<tr>
<td>Pure DNA</td>
<td>Presence of PCR inhibitors</td>
<td>Radstrom, Weissensteiner</td>
</tr>
</tbody>
</table>

* Overview chapters.

## TABLE 2
Mathematical Formulas

<table>
<thead>
<tr>
<th>Aspect of PCR</th>
<th>Chapter</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect of cosolutes on PCR product yield</td>
<td>Chakrabarti (and Reference 14 therein)</td>
<td>Thermodynamic</td>
</tr>
<tr>
<td>Effect of initial target copy and cycle number on PCR product yield</td>
<td>Vu et al.</td>
<td>Statistical</td>
</tr>
<tr>
<td>Equilibrium concentrations of product strand annealing</td>
<td>Ji and Cai</td>
<td>Mass action law</td>
</tr>
<tr>
<td>Number of primer binding sites (cDNA)</td>
<td>Warke et al.</td>
<td>Empirical, statistical</td>
</tr>
<tr>
<td>Number of primer binding sites (genomic DNA)</td>
<td>Remm et al.</td>
<td>Empirical</td>
</tr>
<tr>
<td>Sampling of limiting target copy numbers</td>
<td>Dear</td>
<td>Statistical</td>
</tr>
<tr>
<td>Efficiency of primer extension</td>
<td>Weissensteiner</td>
<td>Empirical</td>
</tr>
<tr>
<td>Experimental design</td>
<td>Weissensteiner</td>
<td>Statistical</td>
</tr>
</tbody>
</table>
REFERENCES


ACKNOWLEDGMENTS

I am grateful to Hugh and Annette Griffin for inviting me to edit this update of their successful first edition of PCR Technology: Current Innovations and to Erika Dery, Project Coordinator at CRC Press, for her patience and continuing support. I would like to thank the authors, with whom it was a pleasure to work on this project. These acknowledgments would not be complete, however, without a special thanks to my family who had to put up with many “lost” evenings and weekends during the preparation of this book.
The Editors

**Thomas Weissensteiner, Ph.D.** studied Biochemistry at Hannover University, Germany, and earned an M.Sc. degree from the Max Planck Institute for Molecular Genetics (Berlin, Germany) in 1989. He worked on the genetic epidemiology of human spondyloarthropathy during his Ph.D. studies at the Deutsche Rheumaforschungszentrum in Berlin and Guy’s Hospital in London, U.K. There he discovered the PCR-enhancing properties of betaine (first published as a conference abstract in 1995) and has since kept a special interest in PCR methodology. He earned a Ph.D. in biochemistry from the Freie Universität Berlin in 1997 followed by postdoctoral research on transgenic animal models for autoimmune joint disease at the Edward Jenner Institute (Compton, U.K.). He has been editor of *Trends in Immunology*, a monthly review journal, since 2001.

**Hugh Griffin, Ph.D.** earned his B.A. (mod) degree in microbiology in 1983 and his Ph.D. in microbiology in 1986 from Trinity College, University of Dublin. He has worked at Washington University, St. Louis, Missouri, the Babraham Institute, Cambridge, U.K., the Institute for Animal Health, Huntingdon, U.K., and the Institute of Food Research, Norwich, U.K. Areas of interest besides PCR include molecular biology of lactic acid bacteria, probiotics, DNA sequence analysis, bioinformatics, and antimicrobial resistance. He currently works at Horizon Scientific Press, Wymondham, U.K.

**Annette Griffin, Ph.D.** graduated in 1982 from University College Cork with a B.Sc. (hons) degree in biochemistry. She earned a Ph.D. in 1986 from Trinity College, University of Dublin. She has worked at the Institute for Animal Health, Huntingdon, U.K., the John Innes Institute, Norwich, U.K., and the Institute of Food Research, Norwich, U.K. Her areas of interest besides PCR include Semliki Forest virus, infectious laryngotracheitis virus, and bacterial exopolysaccharide biosynthesis. She currently works at Horizon Scientific Press, Wymondham, U.K.
Contributors

Tobias Allander
Department of Clinical Microbiology
Karolinska Hospital
Stockholm, Sweden

Catherine Arnold
Genomics, Proteomics and Bioinformatics Unit
Central Public Health Laboratory
London, UK

Mercedes G. Bermudez
Institute for Reproductive Medicine and Science
St. Barnabas Medical Center
West Orange, New Jersey

Robert F. Bonner
Laboratory of Integrative and Medical Biophysics
National Institute of Child Health and Human Development
Bethesda, Maryland

Irene Bosch
Center for Infectious Disease and Vaccine Research and Department of Medicine
University of Massachusetts Medical School
Worcester, Massachusetts

Jens Bukh
Hepatitis Viruses Section
Laboratory of Infectious Diseases
National Institute of Allergy and Infectious Diseases
Bethesda, Maryland

Stephen A. Bustin
Centre for Academic Surgery
Barts and the London, Queen Mary’s School of Medicine and Dentistry
University of London
London, UK

John M. Butler
Biotechnology Division
National Institute of Standards and Technology
Gaithersburg, Maryland

Li Cai
ViaGen, Inc.
Austin, Texas

Raj Chakrabarti
Department of Chemistry
Massachusetts Institute of Technology
Cambridge, Massachusetts

Guo Jun Chen
F. Hoffmann-La Roche AG
Basel, Switzerland

Eric A. Davidson
Department of Chemistry and Biochemistry
Institute for Cell and Molecular Biology
University of Texas
Austin, Texas

Paul H. Dear
MRC Laboratory of Molecular Biology
PNAC Biotech Division
Cambridge, UK

Andrew D. Ellington
Department of Chemistry and Biochemistry
Institute for Cell and Molecular Biology
University of Texas
Austin, Texas

Suzanne U. Emerson
Molecular Hepatitis Section
Laboratory of Infectious Diseases
National Institute of Allergy and Infectious Diseases
Bethesda, Maryland
Sarah B. Faulkner  
Department of Chemistry and Biochemistry  
Institute for Cell and Molecular Biology  
University of Texas  
Austin, Texas

Giordana Feriotto  
Biotechnology Center  
Ferrara University  
Ferrara, Italy

Bernard Fromenty  
INSERM U-481  
Faculté de Médecine Xavier Bichat  
Paris, France

Marcia F. Fournier  
Life Science Division  
Lawrence Berkeley National Laboratory  
Berkeley, California

Roberto Gambari  
Department of Biochemistry and Molecular Biology  
University of Ferrara  
Ferrara, Italy

Jorge A. Garcés  
Genzyme Genetics  
Framingham, Massachusetts

Ray H. Gavin  
Department of Biology  
Brooklyn College, City University of New York  
Brooklyn, New York

Ivo Glynne Gut  
Centre National de Génotypage  
Evry Cedex, France

Baback Gharizadeh  
Department of Biotechnology  
AlbaNova University Center, SCFAB  
Royal Institute of Technology  
Stockholm, Sweden

Sinuhe Hahn  
Laboratory for Prenatal Medicine  
Department of Research  
University Women’s Hospital  
University of Basel  
Basel, Switzerland

Wolfgang Holzgreve  
Laboratory for Prenatal Medicine  
Department of Research  
University Women’s Hospital  
University of Basel  
Basel, Switzerland

Eric Ka-Wai Hui  
Department of Microbiology, Immunology, and Molecular Genetics  
University of California Los Angeles School of Medicine  
Los Angeles, California

Wan Ji  
ViaGen, Inc.  
Austin, Texas

Karen M. Kelly  
CuraGen Corporation  
New Haven, Connecticut

Rickard Knutsson  
MedProbe  
Lund, Sweden

Anelia I. Kraltcheva  
Department of Biology and Biochemistry and Institute of Molecular Biology  
University of Houston  
Houston, Texas

Ramón Kucharzak  
Centre National de Génotypage  
Evry, France
Manfred Kuepper
Universität des Saarlandes
Genetik
Saarbrücken, Germany

Ants Kurg
Institute of Molecular and Cell Biology
University of Tartu and EGeen, Inc.
Estonian Biocentre
Tartu, Estonia

Doris Lechner
Centre National de Génotypage
Evry, France

Ying Li
Laboratory for Prenatal Medicine
Department of Research
University Women’s Hospital
University of Basel
Basel, Switzerland

Szecheng J. Lo
Department of Life Sciences
Graduate Institute of Basic Medical Sciences
School of Medicine
Chang Gung University
Kwei-Shan, Taiwan

Charlotta Löfström
AnalyCen Nordic AB
Linköping, Sweden

Timothy A. Lohret
CuraGen Corporation
New Haven, Connecticut

Maria Lövenklev
Applied Microbiology
Lund Institute of Technology
Lund University
Lund, Sweden

Osman El-Maarri
Institute of Experimental Haematology
and Transfusion Medicine
Bonn, Germany

Rose G. Mage
Molecular Immunogenetics Section
Laboratory of Immunology
National Institute of Allergy
and Infectious Diseases
National Institutes of Health
Bethesda, Maryland

David Magee
Life Science Division
Lawrence Berkeley National Laboratory
Berkeley, California

Heather Melichar
Center for Infectious Disease
and Vaccine Research
Department of Medicine
University of Massachusetts Medical School
Worcester, Massachusetts

Kirsten Mennicke
Department of Human Paediatrics
University of Lübeck
Lübeck, Germany

Jiri Mestecky
Department of Microbiology
University of Alabama
Birmingham, Alabama

Andres Metspalu
Institute of Molecular and Cell Biology
University of Tartu and EGeen, Inc.
Estonian Biocentre
Tartu, Estonia

Kalim U. Mir
Wellcome Trust Centre for Human Genetics
Oxford, UK

Kyl V. Myrick
Harvard University
Department of Molecular and Cellular Biology
Cambridge, Massachusetts

Vu Q. Nguyen
The Kite
Prague, Czech Republic
Tommy Nordström  
Department of Biotechnology  
AlbaNova University Center, SCFAB  
Royal Institute of Technology  
Stockholm, Sweden  

Gerard J. Nuovo  
Department of Pathology  
Ohio State University Medical Center  
Columbus, Ohio  

Pål Nyrén  
Department of Biotechnology  
AlbaNova University Center, SCFAB  
Royal Institute of Technology  
Stockholm, Sweden  

Harold Obiakor  
Molecular Immunogenetics Section  
Laboratory of Immunology  
National Institute of Allergy  
and Infectious Diseases  
Bethesda, Maryland  

Johannes Oldenburg  
Institute of Transfusion Medicine  
and Immune Haematology  
Frankfurt, Germany  

Sven Opitz  
Department of Human Genetics  
University of Lübeck  
Lübeck, Germany  

Xinghua Pan  
Molecular Staging, Inc.  
New Haven, Connecticut  

Dominique Pessayre  
INSERM U-481  
Faculté de Médecine Xavier Bichat  
Paris, France  

Robert H. Purcell  
Hepatitis Viruses Section  
Laboratory of Infectious Diseases  
National Institute of Allergy  
and Infectious Diseases  
Bethesda, Maryland  

Jiannis Ragoussis  
The Wellcome Trust Centre  
for Human Genetics  
Oxford, UK  

Peter Rådström  
Applied Microbiology  
Lund Institute of Technology  
Lund University  
Lund, Sweden  

Maido Remm  
Institute of Molecular and Cell Biology  
University of Tartu and EGeen, Inc.  
Estonian Biocentre  
Tartu, Estonia  

Michael W. Russell  
Department of Microbiology  
University of Buffalo  
Buffalo, New York  

Philippe Schmitt  
Qbiogene  
Illkirch, France  

Markus Schuelke  
Department of Neuropediatrics  
Charité, Campus Virchow University Hospital  
Humboldt University  
Berlin, Germany  

Eberhard Schwinger  
Department of Human Genetics  
University of Lübeck  
Lübeck, Germany  

Sten Stemme  
Cardiovascular Research Unit  
Center for Molecular Medicine  
Karolinska Institute  
Stockholm, Sweden
Veronika Stemme  
Cardiovascular Research Unit  
Center for Molecular Medicine  
Karolinska Institute  
Stockholm, Sweden

Anne C. Stone  
Department of Anthropology  
Arizona State University  
Tempe, Arizona

Heather A. Syrett  
Department of Chemistry and Biochemistry  
Institute for Cell and Molecular Biology  
University of Texas  
Austin, Texas

Yerramilli V.B.K. Subrahmanyam  
Artesian Therapeutics Inc.  
Gaithersburg, Maryland

Jörg Tost  
Centre National de Génotypage  
Evry, France

Serge Troubetzkoy  
Centre de Physique Théorique  
Institute de Mathématiques de Luminy  
Marseille, France

Peter M. Vallone  
Biotechnology Division  
National Institute of Standards and Technology  
Gaithersburg, Maryland

Huong L. Vu  
Department of Microbiology  
University of Alabama at Birmingham  
Birmingham, Alabama

Jörn Walter  
Universität des Saarlandes  
Genetik  
Saarbrücken, Germany

Rajas V. Warke  
Center for Infectious Disease and Vaccine Research and Department of Medicine  
University of Massachusetts Medical School  
Worcester, Massachusetts

Thomas Weissensteiner  
Elsevier  
Trends in Immunology  
London, UK

Sherman M. Weissman  
Department of Genetics  
Yale University School of Medicine  
New Haven, Connecticut

Dagan Wells  
Institute for Reproductive Medicine and Science  
St. Barnabas Medical Center  
West Orange, New Jersey

Selwyn A. Williams  
Department of Biology  
Brooklyn College, City University of New York  
Brooklyn, New York

Petra Wolffs  
Applied Microbiology  
Lund Institute of Technology  
Lund University  
Lund, Sweden

Xiao Yan Zhong  
Laboratory for Prenatal Medicine  
Department of Research  
University Women’s Hospital  
University of Basel  
Basel, Switzerland
Contents

SECTION I  General Considerations

Template Isolation and Preparation

Chapter 1  Extraction and Amplification of Ancient DNA............................................................5
Anne C. Stone

Chapter 2  Long PCR from Damaged DNA................................................................................11
Bernard Fromenty, Philippe Schmitt, and Dominique Pessayre

Chapter 3  Quantitative mRNA Analysis in Small Cell Samples by RT-PCR
and Flow Cytometry...................................................................................................23
Veronika Stemme and Sten Stemme

Chapter 4  Microdissection of Single or Small Numbers of Cells for Analyses
of DNA and RNA by PCR....................................................................................29
Rose G. Mage, Robert F. Bonner, and Harold Obiakor

Chapter 5  Pre-PCR Processing Strategies................................................................................37
Peter Rådström, Maria Lövenklev, Petra Wölffs, Charlotta Löfström, and Rickard Knutsson

PCR Optimization

Chapter 6  Novel PCR-Enhancing Compounds and Their Modes of Action.........................51
Raj Chakrabarti

Chapter 7  Optimizing PCR with the Aid of Experimental Design......................................65
Thomas Weissensteiner

Labeling and Cloning

Chapter 8  Economic Fluorescent Labeling of PCR Products for Microsatellite-
and Single-Stranded Conformation Polymorphism (SSCP) Analysis.........................79
Markus Schuelke

Chapter 9  Universal Restriction Site-Free Cloning Method Using Chimeric Primers........87
Guo Jun Chen
SECTION II  Analytical Applications

Detection of Nucleic Acid Variation

Chapter 10  High-Throughput Genotyping and the PCR .........................................................97
Kalim U. Mir and Jiannis Ragoussis

Chapter 11  High-Throughput Genetic Analysis through Multiplexed PCR
and Multicapillary Electrophoresis .............................................................................111
John M. Butler and Peter M. Vallone

Chapter 12  The GOOD Assay: A Purification-Free Assay for Genotyping by MALDI
Mass Spectrometry ...................................................................................................121
Jörg Tost, Ramón Kucharzak, Doris Lechner, and Ivo Glynne Gut

Chapter 13  Primer Design for Large-Scale Multiplex PCR and Arrayed Primer
Extension (APEX) ....................................................................................................131
Maido Remm, Ants Kurg, and Andres Metspalu

Chapter 14  Surface Plasmon Resonance Based Biosensor Technology for Real-Time
Detection of PCR Products .......................................................................................141
Giordana Feriotto and Roberto Gambari

Chapter 15  Fluorescent Amplified Fragment Length Polymorphism (FAFLP) Genotyping.....155
Catherine Arnold

Chapter 16  Global Analysis of DNA Allelic Variations (GADAV) by Speciﬁc
Enrichment of Mismatches and Selective Ampliﬁcation of Heterohybrids ............163
Xinghua Pan and Sherman M. Weissman

Chapter 17  Quantitative DNA-Methylation Analysis by the Bisulﬁte Conversion Method.....175
Osman El-Maarri, Manfred Kuepper, Johannes Oldenburg, and Jörn Walter

Chapter 18  Pyrosequencing Technology .............................................................................187
Pål Nyrén, Tommy Nordström, and Baback Gharizadeh

Chapter 19  Octamer-Primed Cycle Sequencing ................................................................197
Anelia I. Kralcheva and Susan H. Hardin

Chapter 20  In Situ Amplification of cDNA ........................................................................205
Gerard J. Nuovo

Chapter 21  Detection of a Single-Copy Sequence on Human Metaphase by Cyclic PRINS.....215
Kirsten Mennicke, Sven Opitz, and Eberhard Schwinger
**Nucleic Acid Quantitation**

**Chapter 22** Meaningful Quantification of mRNA Using Real-Time PCR ..........................225
Stephen A. Bustin

**Chapter 23** Absolute Quantification of Specific Nucleic Acids by RT-PCR Using a Nonlinear Mathematical Model for Data Analysis ..................................235
Huong L. Vu, Serge Troubetzkoy, Vu Q. Nguyen, Michael W. Russell, and Jiri Mestecky

**Chapter 24** Single-Molecule PCR – Basic Protocols and Applications ..............................245
Paul H. Dear

**Chapter 25** Gene Expression and Discovery: Application of a Genome-Wide Approach ......259
Timothy A. Lohret and Karen M. Kelly

**Chapter 26** The Detection of Differential Gene Expression Changes Using SAGE (Serial Analysis of Gene Expression) ...............................................................267
Jill Powell

**Chapter 27** Applications of Differential Display in Infectious Diseases .................................277
Irene Bosch, Rajas V. Warke, Marcia F. Fournier, David Magee, and Heather Melichar

**Chapter 28** Restriction Endonucleolytic Analysis of Differentially Expressed Sequences: READS .................................................................287
Yerramilli V. B. K. Subrahmanyam and Sherman M. Weissman

**Chapter 29** Suppression Subtractive Hybridization PCR .....................................................299
Wan Ji and Li Cai

**SECTION III Preparative Application**

**Amplification of Unknown Sequences**

**Chapter 30** PCR-Based Techniques for Cloning an Unknown DNA Fragment Adjacent to a Known Integrant .................................................................313
Eric Ka-Wai Hui and Szecheng J. Lo

**Chapter 31** Chromosome Walking by Inverse PCR ..........................................................323
Jorge A. Garcés, Selwyn A. Williams, and Ray H. Gavin

**Chapter 32** Long Universal Fast Walking ........................................................................329
Kyl V. Myrick
Chapter 33  Cloning of Unknown Virus Sequences by DNase Treatment and Sequence-Independent Single Primer Amplification ........................................337
Tobias Allander, Suzanne U. Emerson, Robert H. Purcell, and Jens Bukh

Amplification of Minute Amounts of Nucleic Acids

Chapter 34  Whole Genome Amplification from Single Cells and Minute DNA Samples ......349
Dagan Wells and Mercedes G. Bermudez

Chapter 35  Primer Extension Preamplification (PEP) of Single Cells: Efficiency and Bias .....361
Xiao Yan Zhong, Ying Li, Wolfgang Holzgreve, and Sinuhe Hahn

In Vitro Evolution

Chapter 36  In Vitro Evolution through PCR-Mediated Mutagenesis and Recombination ......371
Sarah B. Faulkner, Heather A. Syrett, Eric A. Davidson, and Andrew D. Ellington

Index ..............................................................................................................................................381
Section I

General Considerations
Template Isolation and Preparation
1 Extraction and Amplification of Ancient DNA

Anne C. Stone

CONTENTS

I. Introduction ................................................................. 5
II. Materials and Methods .................................................. 6
  A. Sample Preparation ................................................ 6
  B. Phenol Chloroform Extraction Protocol ....................... 6
  C. Silica and Guanidinium Thiocyanate Extraction Protocol .. 6
  D. PCR Amplification ................................................ 7
III. Discussion ................................................................ 7
References ........................................................................ 9

I. INTRODUCTION

Ancient DNA can be extracted from cellular remains such as bone, tooth roots, tissue (preserved under water, frozen, or dried), coprolites, seeds, and other plant materials, and thus, the questions that can be addressed are wide-ranging from the history of the domestication of plants and animals to the relationships between individuals at an archaeological site or the relationships between extant and extinct species. Ancient DNA analyses, however, are fraught with difficulty both because of the small quantity and damaged nature of the DNA and because of problems with contamination from modern sources. These difficulties require special facilities and techniques to enhance the probability of recovery and to assure the authenticity of the results. The work should be performed in a laboratory that is located separately from the main laboratory where modern DNA and post-PCR products are analyzed and stored, and equipment and reagents should be dedicated to ancient DNA work. Reagents must be tested for contamination by including both extraction and PCR “blanks” which are subject to the same protocols as the primary samples. All results also must be verified by multiple independent extractions and analyses. Finally, the results should make sense (i.e., the sequences should not match one from the investigator’s genome or match that of a cow sequence when a sloth sequence is expected).

Ancient DNA work requires much care and patience, and, because of the necessity of multiple independent analyses, it is often expensive. As a result, the questions asked and the probability of actually getting sufficient data to answer the questions posed should be carefully evaluated before undertaking an ancient DNA research project. Two main protocols have been used to extract DNA from ancient or degraded remains. The first is a proteinase K digestion followed by a phenol/chloroform extraction. The second primary technique is the silica/guanidine thiocyanate technique which will be discussed in detail here.
II. MATERIALS AND METHODS

A. SAMPLE PREPARATION

The first step in the extraction protocol is to prepare the sample by removing any surface contamination from previous handling of the material by excavators, archaeologists, or osteologists, as well as laboratory workers. Surface contamination is removed by cutting or grinding away the exposed layers, by irradiating the surface with 254-nm ultraviolet (UV) light for 30 min, or by soaking the material for 10 min in 1 N hydrochloric acid or 10% sodium hypochlorite (10% vol/vol). Acid or sodium hypochlorite should not be used in cases where either can penetrate deeply into the material. Gloves, a face mask, a face shield, and a laboratory coat should be worn during sample preparation. For bones, the outer layer can be removed with a rotary tool with a grinding stone bit (Sears Craftsman®). When teeth are used, the whole tooth is soaked in 10% sodium hypochlorite for approximately 10 min and then rinsed with 70% ethanol. The tooth cap is then removed with the rotary tool (using the cut-off wheel accessory) at the enamel–dentin boundary and the root used for DNA extraction. After a bone or tooth root sample is cleaned, it is ground to dust in a bone mill (SPEX, Inc.) or using a steel nut and bolt, while soft tissue samples are cut into fine pieces with a sterile scalpel. This increases the surface area, enhancing the release of DNA during the extraction. Vials used with the bone mill should be washed with H₂O, soaked in 1 N HCl or 10% sodium hypochlorite for at least 15 min, rinsed with double distilled water (ddH₂O) and UV irradiated between uses. After the samples are ground into dust, they should be stored in sterile 15-ml tubes at room temperature or in a refrigerator or freezer until used in the extraction.

B. PHENOL CHLOROFORM EXTRACTION PROTOCOL

Digestion with proteinase K breaks up the proteins in the tissue and releases the DNA, while the phenol chloroform extraction separates the DNA from the proteins in the solution. This method is standard in biology and results in a high yield of DNA. For bone, some investigators precede the phenol/chloroform method with a soak in ethylene diaminetetra-acetic acid (EDTA) buffer to decalcify the bone sample.⁶

C. SILICA AND GUANIDINIUM THIOCYANATE EXTRACTION PROTOCOL

The silica and guanidinium thiocyanate protocol for ancient DNA extraction takes advantage of the nucleic acid binding properties of silica in the presence of the chaotropic agent guanidinium thiocyanate.³,⁴ This protocol requires the preparation of size fractionated (nonfumed) silica and two buffers, L6 and L2.

1. Size Fractionated Silica: Add 12 g of silicon dioxide and ddH₂O to a total volume of 100 ml in a glass-graduated cylinder. Vortex and then sediment at unit gravity for 24 h at room temperature. Remove 86 ml of the supernatant and add ddH₂O to 100 ml. Vortex to resuspend the silica pellet. Sediment again for 5 h and then remove 88 ml of the supernatant. Add 120 μl of HCl (32% wt/vol) to adjust to pH 2. Pipet the silica into 1.5 ml tubes and store in the refrigerator in the dark (stable in the dark for at least 3 months). If the pH of the silica rises above 2, the silica will not work.

2. Buffer L2: Working under a hood, add 20 ml of 0.1 M Tris hydrochloride pH 7.4 to 24 g Guanidinium thiocyanate (GuSCN) in a 50-ml falcon tube. Heat to 60–65°C to facilitate dissolution (or microwave carefully on “high” for 15 sec, shake, and then microwave for an additional 15 sec).

3. Buffer L6: Prepare buffer L2 as described above for L6. Then, add 4.4 ml of 0.2 M EDTA (adjusted with NaOH to pH 8) and 0.5 ml Triton X-100. Finally, add 1.5 ml fractionated silica (vortex prior to pipetting) to bind any contaminating DNA, vortex, centrifuge, and
divide the supernatant into two aliquots. Both L2 and L6 buffers are stable for at least 1 month at room temperature in the dark.

4. DNA extraction: Add 0.25 to 0.5 g of powdered bone or tooth root and 820 μl L6 to a 1.5-ml tube. A blank containing reagents only should be included in the extraction procedure with each bone or tooth root sample. Vortex and incubate in a heating block at 60°C for 2 h with occasional vortexing. Centrifuge tubes and take 500 μl of the supernatant to a new tube containing a mixture of 500 μl of L6 and 40 μl of silica (vortex prior to pipetting) and vortex well. Leave for 15 min to allow the DNA to bind to the silica and then centrifuge in a bench-top centrifuge at 11,750 RCF (relative centrifugal force) for at least 15 sec. Remove the supernatant and wash pellet twice with L2, twice with 70% ethanol, and once with acetone (add 1 ml each time, resuspend silica completely by vortexing, centrifuge for 15 sec and discard the supernatant). Dry the pellet at 56°C with open lids (lightly covered with aluminum foil) for 10 min. Add 65 μl ddH₂O, vortex to resuspend the silica, leave at 56°C for 10 min, centrifuge and pipet supernatant into a new tube while removing as little silica as possible. Reextract the pellet with another 65 μl of ddH₂O, adding the supernatant to the same tube. Use 5 μl of supernatant for PCR (centrifuge the sample first to sediment any silica transferred with the supernatant since it will inhibit PCR).

D. PCR Amplification

After extraction, PCR is used to copy the DNA fragment of interest millions of times so that there is sufficient DNA for analysis. For ancient DNA, a number of special steps are necessary to ensure successful amplification. First, because ancient DNA is normally degraded to fragments 100 to 200 bp in length and found in low amounts, primers must be designed to amplify small fragments with high efficiency and specificity. The sensitivity of the assay should be tested using a series of control DNA dilutions. Second, adding bovine serum albumin (BSA) in the PCR to bind inhibitors (50 μg in a 50 μl reaction) greatly aids amplification. Third, a hot start PCR should be performed. For a hot start PCR, the Taq polymerase is separated physically or chemically from the other reagents, in particular the primers, so that unspecific priming at low temperatures cannot occur (e.g., AmpliTaq Gold). Finally, the number of PCR cycles is typically increased to 40. Each PCR should include both extraction and PCR blanks to check for contamination. After PCR, up to 10 μl of the PCR products are loaded on an agarose gel for visualization. Often, several PCR attempts are necessary to obtain a product. After PCR, the DNA can be analyzed by cutting the DNA with restriction enzymes that cleave the DNA at sequence-specific sites, or by DNA sequencing, either directly or after cloning. Prior to direct sequencing or cloning, it is useful to reamplify the PCR product so that more DNA is produced. To do this, run the PCR product on a low-melting agarose gel such as NuSieve (FMC), cut out the product band, and place it in a tube with 100 μl of ddH₂O. Place this tube in a 65°C water bath to melt the gel plug and then reamplify from 3 μl of this solution under the same conditions as the previous PCR, but with only 25 cycles and the annealing temperature raised by 2°C.

III. DISCUSSION

Successful extraction of ancient DNA depends on the conditions of preservation both in and out of the ground (success rates range from 0 to 90%). Each extraction method has particular advantages and disadvantages for ancient DNA. One disadvantage of the phenol/chloroform protocol for ancient DNA work has been the coextraction of inhibitors such as fulvic acids from the soil that make it difficult to amplify and analyze the DNA. The other primary method, the silica/guanidine thiocyanate technique, has the advantage of removing the inhibitors. In addition, it is simple and fast; however, the DNA yield may not be as high. Other methods for extracting DNA from ancient remains include using silica-based spin columns, cetyltrimethylammonium (CTAB) buffer, chelex, or using a combination of the two primary methods reported here.
Chelex is a resin that binds to impurities but not to DNA which can then be recovered from the solution. However, DNA prepared by this method tends to degrade over time, and therefore must be analyzed quickly. Some researchers also have added extra steps to their extraction method in order to repair DNA damage, such as crosslinking (where the DNA molecule chemically links to other DNA molecules or to itself), oxidative damage,13 and other forms of single strand breaks (see Chapter 2 by Bernard Fromenty on exonuclease-III pretreatment).

Success also depends on the genetic locus examined. Most ancient DNA research has targeted mitochondrial DNA (mtDNA). Unlike nuclear DNA, mtDNA is present in almost 1000 copies per cell. The high copy number means that mtDNA is more likely to survive over time than nuclear DNA. However, attempts to examine nuclear DNA loci, including Y chromosome sequences for sex identification, nuclear genes for phylogenetic analysis, and short tandem repeat (STR) loci for determining relatedness between individuals, also have been successful. Finally, a few studies have investigated bacterial or viral DNA from ancient materials.

Yield and quality of ancient DNA are strongly affected by the environment where samples are found and by the treatment of samples after they are recovered. In general, cooler temperatures, neutral or slightly alkaline pH of the soil, and dry environments are best for DNA preservation. To date, cave and permafrost environments have yielded the oldest reliable DNA results. Microenvironments within a burial or excavation site can also influence DNA recovery by affecting the rate of oxidative or hydrolytic DNA damage which can inhibit PCR or cause misincorporation of nucleotides. In particular, oxidative damage can cause strand breakage, while deamination of cytosine, one type of hydrolytic damage, appears to be fairly common in ancient DNA and can lead to substitution artifacts.14–16 Other environmental factors affecting DNA recovery and analysis include substances that may be coextracted with the DNA and that inhibit PCR. These substances include fulvic acids, which are breakdown products of organic soils, and Maillard reaction products, which are produced during the initial decay of organic matter.

After excavation, preservatives such as formaldehyde, alcohol, shellac, gamma radiation, glue, and tanning agents often have been used to treat samples. Many of these destroy or degrade the remaining DNA.17 Moreover, these preservatives can introduce contaminants from the preparer, from previous samples that were treated with the preservative, or from the preservative itself, many of which were traditionally made from plant or animal materials.18 For samples that have been treated with preservatives, using bovine serum albumin (BSA) in the PCR to bind some inhibiting molecules, removing surface treatments, and modifying the extraction method may aid in ancient DNA recovery. Common PCR inhibitors and methods for overcoming them are reviewed by Peter Rådeström et al. in Chapter 5. In general, withholding preservative treatment during or after excavation for samples that will be subject to DNA analysis is best.

DNA recovery is also dependent on the type of material used for extraction. Tooth roots are generally the best source of ancient DNA. If teeth are not available, the quality and quantity of DNA isolated from bone is better than that from soft tissue. Hydroxyapatite, which forms the framework for bones and tooth roots, may be the reason for the difference between preservation in hard and soft tissues. Hydroxyapatite binds DNA and may buffer it from damage and degradation.

The next best sources are peripheral tissues from mummified individuals. These tissues are more likely to desiccate rapidly after death and thus their DNA might escape extensive degradation from lytic enzymes.

Regardless of the protocols used to analyze ancient human DNA, the greatest concern is contamination. In addition to the precautions mentioned previously, several steps help to identify and limit contamination. First, aliquot all reagents into smaller tubes so that these can be quickly used up and thrown out. Second, test all new reagents using three PCR blanks so that low-level contamination is apparent. Even when analyzing nuclear loci, use test reagents amplifying mtDNA loci as these are generally more sensitive indicators of contamination.19 Third, a PCR enclosure (without blowing air and which allows only the arms to fit into the space) helps to limit contamination.
REFERENCES

I. INTRODUCTION

Several types of DNA lesions can hamper or block the progress of DNA polymerase during PCR amplification. Although the most common blocking lesions are DNA strand breaks and apurinic/apyrimidinic (AP) sites (also called abasic sites), other blocking lesions may exist. The larger the DNA sequence, the greater the probability that one or several blocking lesions will be present on each DNA molecule. Thus, whereas DNA must be extensively degraded to partially hamper amplification of a short DNA fragment, even mild DNA damage can totally prevent Long PCR amplification. Indeed, while DNA samples had to be heated at 99°C for more than 10 min to partly decrease amplification of a 316-bp mitochondrial DNA (mtDNA) sequence, heating DNA for only 2 or 3 min prevented the amplification of a 8600-bp mtDNA fragment.

In the absence of genotoxic conditions or treatments, blocking DNA lesions are uncommon in living organisms. DNA strand breaks are absent or extremely rare, as these lesions are incompatible with cell life, and only 10 to 30 abasic sites are found per 10^6 nucleotides in normal rat or human tissues. These low physiological levels of blocking lesions are unlikely to significantly hamper Long PCR. However, the factors listed below can lead to more extensive DNA damage.

Postmortem accumulation of DNA blocking lesions is a major issue in archeological and forensic science. In archeological DNA samples that are 13,000- to more than 50,000-years-old, the probability of finding blocking lesions is high.
old, extensive DNA lesions may even prevent the amplification of a short (140-bp) mtDNA fragment.  

DNA extraction also causes DNA damage, which must be minimized for successful Long PCR. Although the phenol extraction method is not the best technique to recover high molecular weight DNA, it nevertheless allows amplification of 10,000- to 20,000-bp DNA sequences. Extracting DNA in the presence of 2,2,6,6-tetramethylpiperidinoxyl (TEMPO) has been suggested, although the effects of this radical-trapping agent on long PCR amplification have not yet been assessed.

Conservation and storage of tissues and DNA are also important. Fixatives such as Carnoy’s, Zenker’s, or Bouin’s fluid alter DNA and prevent amplification of DNA fragments longer than 1000 bp from paraffin-embedded tissues. Even conserving DNA for several years in a freezer can damage DNA and hamper long PCR amplification. Further investigations are needed to find conservation/storage conditions to preserve the integrity of DNA.

The PCR reaction itself can generate blocking DNA lesions. The high temperatures that are used in PCR reactions favor depurination, which occurs four times faster with single-stranded DNA than double-stranded DNA. Therefore, a short DNA denaturation step (temperature above 90°C) is recommended for long PCR amplification. Depurination is also enhanced at low pH. Since the pH of some PCR buffers (e.g., Tris buffer) can decrease at high temperatures, this low pH may further depurinate DNA. Not only do AP sites hamper the progress of polymerases, they also weaken the DNA chain, and can trigger DNA cleavage through a \( \beta \)-elimination process. Although DNA breakage is increased by the presence of Mg\(^{2+}\), this cation cannot be omitted from long PCR media since Mg\(^{2+}\) is mandatory for DNA polymerase activity. DNA breakage is also increased by polyamines (e.g., spermine, spermidine, putrescine), which can be omitted from long PCR media. Even though spermidine facilitates PCR amplification of short DNA fragments, a high concentration (50 mM) of putrescine decreased the amplification of a long DNA sequence (ca. 8.6-kb), suggesting that polyamines are best omitted for long PCR amplifications.

Depending on the goal of the PCR, the impairment of long PCR by DNA lesions can be either a useful research tool or an impediment that must be overcome:

1. Because DNA damage hampers the amplification of long DNA sequences, this is a convenient tool to detect DNA lesions on DNA targets. A method to assess mtDNA damage with a 4-primer long PCR protocol is presented in Section II of this chapter.
2. On the other hand, the impairment of long PCR amplification by DNA blocking lesions can seriously hinder amplification of large DNA sequences. During studies to ameliorate the long PCR technology as a tool to detect DNA damage, we discovered that exonuclease III (a major repair enzyme of *Escherichia coli*) markedly enhances long PCR amplification of damaged DNA templates. Protocols that take advantage of this effect are provided in Section III of this chapter.

## II. USE OF LONG PCR TO DETECT MTDNA DAMAGE

In the past few years, long PCR techniques have been successfully used in toxicology to detect the mtDNA damage induced by oxidative stress in diverse experimental models. This long PCR approach has several advantages, including rapidity, sensitivity, specificity, and easy quantification (see the following text for details).

### A. DNA ISOLATION, MATERIALS, AND PCR INSTRUMENTATION

Total DNA from mouse liver was isolated using QIAGEN\textsuperscript{\textregistered} Genomic-tip 100/G columns (QIAGEN, Hilden, Germany). Desalted primers were obtained from Life Technologies (Cergy-Pontoise, France) and resuspended in sterile distilled water to obtain 100 pmol/\textmu l stock solutions that were diluted 10-fold when needed. PCR grade dNTPs (sodium salts) were purchased from Roche
Long PCR from Damaged DNA

Molecular Biochemicals (Meylan, France). Each dNTP solution was diluted 10-fold with sterile distilled water, and stored at –20°C as a 10 mM stock solution. Nusieve® GTG® and Seakem® GTG® agaroses were from FMC BioProducts (Rockland, ME).

Long PCR amplification of mouse mtDNA was performed in 0.2 ml MicroAmp™ reaction tubes (Perkin-Elmer, Courtaboeuf, France) using a RoboCycler Gradient 96 PCR apparatus equipped with a heat cover (Stratagene, Montigny-le-Bretonneux, France). Amplification reactions were performed with the GeneAmp® XL PCR system (Perking-Elmer), involving a mixture of rTth DNA polymerase and Vent® DNA polymerase. The latter polymerase exhibits 3’–5’ exonuclease (proofreading) activity. The Perkin Elmer XL buffer is composed of tricine, potassium acetate, glycerol, and DMSO.

B. PCR PROTOCOL

1. Principle of the Four-Primer Long PCR Protocol

A protocol using four primers was designed to assess the presence of blocking lesions on mtDNA. This protocol coamplifies a long (8636-bp) mtDNA fragment and a short (316-bp) mtDNA fragment. As indicated above, random DNA lesions are much more likely to decrease amplification of a long DNA fragment than a short fragment. Therefore, the presence of blocking lesions on the mitochondrial genome is expected to markedly decrease the yield of the 8636-bp PCR product with little effect, or no effect, on the yield of the 316-bp product.

2. PCR Conditions

Forward primer A 5’-CGACAGCTAAGACCCAAACTGGG-3’ and backward primer B 5’-CCCATTTCCTCCATTTGAGC-3’ amplify a 316-bp mtDNA fragment, while forward primer C 5’-TACCTCTGCCAGATCCATTTGAGC-3’ and backward primer D 5’-GAGGCTGATCTTTGATTTGACGGA-3’ amplify an 8636-bp mtDNA fragment (see Reference 23 for mouse mtDNA sequence). The long and short DNA fragments must be nonoverlapping. In order to get specific and efficient amplification of the two DNA targets without spurious bands, we tested in preliminary experiments different combinations of primers and, for the preselected combinations, several molar ratios between the two primer couples. The PCR reactions are performed in a volume of 50 µl, with primers A and B (14 pmol), C and D (40 pmol), total DNA (50 to 200 ng), each dNTP (200 µM), magnesium acetate (1.75 mM), and rTth DNA polymerase XL (1.7 units) in MicroAmp™ reaction tubes (Perkin-Elmer).

PCR reactions are performed as follows:

1. For 10 reactions, prepare master mix A and keep on ice:
   - Sterile distilled water 57 µl
   - 3.3 × GeneAmp® XL PCR buffer 60 µl
   - dNTP (10 mM solutions) 10 µl each
   - Primers A and B (10 pmol/µl) 40 µl each
   - Primers C and D (10 pmol/µl) 14 µl each
   - Magnesium acetate (25 mM solution) 35 µl
2. For 10 reactions, prepare master mix B and keep on ice:
   - Sterile distilled water 91.5 µl
   - 3.3 × GeneAmp® XL PCR buffer 90 µl
   - DNA polymerase XL 8.5 µl
3. Add 30 µl of mix A, 19 µl of mix B, and 1 µl of the DNA sample in each PCR tube.
4. Place the tubes in the thermocycler and start the PCR program (note that with a PCR apparatus that is not equipped with a heat cover, mineral oil is needed). The thermocycler profile includes initial denaturation at 95°C for 2 min, 26 cycles of 95°C for 45 sec, 57°C for 45 sec, and 68°C for 7.5 min, and final extension at 68°C for 7 min.
3. Analysis of PCR Products and Interpretation of Data

PCR products (20 µl) are electrophoresed on ethidium bromide–containing 1.6% agarose gels (0.7% Nusieve® GTG™ agarose plus 0.9% Seakem® GTG™ agarose). After migration, photographs are taken under UV transillumination and scanned to determine the 8636-bp product/316-bp product intensity ratio.

Figure 2.1A shows the PCR products that were obtained with this 4-primer Long PCR protocol, using DNA samples that had been heated at 99°C for 30, 60, 90, or 120 sec (lanes 1 to 4). As DNA is increasingly damaged, the yield of the 8636-bp mtDNA fragment progressively decreases (lanes 1 to 3), to finally disappear in lane 4. In contrast, the 316-bp product remains constant. Selective impairment of the amplification of the long fragment, but not the short fragment, reflects the much higher probability of having DNA lesions that block the progress of the polymerases on a long, rather than a short, DNA stretch. (A) In the absence of exonuclease III (exo 0), the 8636-bp PCR product progressively decreases as DNA damage increases (lanes 1 to 3), to finally disappear in lane 4. In contrast, the 316-bp product remains constant. Selective impairment of the amplification of the long fragment, but not the short fragment, reflects the much higher probability of having DNA lesions that block the progress of the polymerases on a long, rather than a short, DNA stretch. (B) Adding 25 units of exonuclease III to the PCR medium (exo 25 U) markedly increases the yield of the 8636-bp PCR product from mildly or moderately damaged DNA samples (lanes 1 to 3) and restores amplification of the severely damaged DNA sample that was not amplified without exonuclease III (lane 4). M, HindIII- digested phage λ DNA (fragment sizes 23.1, 9.4, 6.6, 4.4, 2.3, 2.0, and 0.56 kb).

The average number (λ) of DNA lesions totally blocking the progress of polymerases per DNA strand can be calculated from Poisson’s equation: λ = –ln A_d/A_c, where A_d and A_c are the amounts...
of PCR product obtained from the damaged DNA template and the control template, respectively.\textsuperscript{18–20} With this 4-primer Long PCR method, the $A_d$ and $A_c$ values are the 8636-bp product/316-bp product intensity ratios obtained with damaged and control DNA templates, respectively.

C. \textbf{POSSIBLE PITFALLS, TROUBLESHOOTING TIPS, AND OTHER REMARKS}

For some DNA samples, the yield of the 316-bp mtDNA fragment can be low or nil. One reason can be complete degradation of mtDNA. This can be assessed by Southern blot hybridization with an mtDNA probe. While moderate oxidative damage to mtDNA decreases the supercoiled and circular mtDNA forms and increases linear mtDNA fragments (full length 16.3-kb mtDNA form and possibly shorter fragments migrating as a smear),\textsuperscript{20} extensive mtDNA damage (induced, for instance, by heating DNA at 99°C for several minutes) suppresses all hybridization signals, as mtDNA is cut into tiny fragments that are no longer recognized by the mtDNA probe.\textsuperscript{5}

Another possible reason for decreased amplification of the short 316-bp mtDNA fragment can be the presence of PCR inhibitors in the reaction medium. A “mixing” experiment can assess this possibility. The DNA sample that is suspected of containing PCR inhibitors is mixed with a second DNA sample that has been successfully amplified in a previous PCR experiment. Impaired amplification of the mixed sample suggests the presence of PCR inhibitors in the first sample. Further DNA extraction can be attempted to get rid of the putative inhibitors.

In some circumstances, mtDNA levels can be markedly reduced while total DNA is unaffected. Very low mtDNA amounts in the PCR reaction can decrease PCR yields. Lowering the amount of total DNA below 10 ng caused a preferential loss of the short mtDNA fragment compared to the long mtDNA fragment, thus increasing the 8636-bp product/316-bp product intensity ratio. This ratio also increased when the number of PCR cycles was decreased. However, when the amount of total DNA was varied between 10 and 250 ng in the PCR reaction, the 8636-bp product/316-bp product intensity ratio remained nearly constant.\textsuperscript{5} Thus, if the amount of total DNA is used as recommended (i.e., 50 to 200 ng of total DNA for our 4-primer Long PCR protocol with primer pairs A-B and C-D), the yield of the short fragment should not decrease except with extremely damaged DNA samples or in the presence of PCR inhibitors, as indicated above.

Finally, it is important to emphasize that setting up a new PCR protocol with four primers in the same reaction is not always easy. Preliminary experiments are necessary to determine the best molar ratio between the two primer couples. Primers must be changed if spurious products are amplified.

III. \textbf{USE OF EXONUCLEASE III TO ENHANCE LONG PCR AMPLIFICATION OF DAMAGED DNA TEMPLATES}

Sometimes DNA has been involuntarily damaged, and one nevertheless wishes to recover long DNA sequences. We recently found that adding \textit{E. coli} exonuclease III to the amplification medium greatly enhanced the yield of long PCR products from damaged DNA templates. We observed this beneficial effect with old DNA samples (DNA samples kept in the freezer for several years) and with DNA samples that were submitted to diverse \textit{in vitro} treatments generating strand breaks and/or AP sites (high temperature, depurination at low pH, or near-UV irradiation). Amplification of products ranging in size from 5 to 15.4 kb from nuclear as well as mtDNA target sequences was also improved.\textsuperscript{4} Importantly, improvement was obtained with different DNA polymerases or their combinations, such as \textit{rTth} DNA polymerase, \textit{rTth} plus \textit{Vent} DNA polymerases, or \textit{Taq} plus \textit{Pwo} DNA polymerases. More recently, we found that addition of exonuclease III also improved the amplification of shorter DNA target sequences with \textit{Taq} and \textit{Tfu} DNA polymerases (see following text).
A. EFFECTS OF EXONUCLEASE III ON THE AMPLIFICATION OF DAMAGED mtDNA TEMPLATES

We used the 4-primer Long PCR protocol described above to amplify mtDNA from total DNA samples that had been heated at 99°C for 30, 60, 90, or 120 sec (lanes 1 to 4, Figure 2.1A and Figure 2.1B). Supplementing the PCR medium with 25 units of *E. coli* exonuclease III (New England Biolabs, Beverly, MA) greatly enhanced amplification of the 8.6-kb fragment (exo+, lanes 1 to 3), and even restored amplification from the most severely damaged DNA sample (exo+, lane 4, Figure 2.1B). Similar beneficial effects were obtained when damaged DNA samples were first incubated at 37°C for 30 min with exonuclease III (in its appropriate buffer) prior to the PCR reaction (data not shown).

High concentrations of exonuclease III (combined with low amounts of the A/B primer couple) prevented amplification of the short (316-bp) mtDNA fragment (Figure 2.1B). This might be due to the destruction of primers A and B, which are present in low amounts (14 pmol) in the standard reaction medium. Amplification of the 316-bp fragment was restored by using less exonuclease III (1 or 5 units instead of 25) in the PCR medium or by preheating exonuclease III, which inhibits its 3'-5' exonuclease activity. These modified procedures did not hamper the beneficial effect of exonuclease III on the amplification of the large, 8.6-kb mtDNA product.

B. EFFECTS OF EXONUCLEASE III ON THE AMPLIFICATION OF DAMAGED NUCLEAR DNA TEMPLATES

1. Extraction of DNA, Materials, and PCR Instrumentation

Total DNA from human blood was isolated using QIAGEN® columns (Blood and Cell Culture DNA Midi Kit). DNA was extracted from freshly drawn blood samples and also from a blood sample that had been kept at −20°C for 1 year but was then thawed due to freezer failure. In some experiments, DNA from fresh blood was voluntarily denatured at 99°C for 60, 90, and 120 sec in the thermocycler.

The dNTP mix (5 mM each) and the primers were from Qbiogene, whereas exonuclease III was purchased from New England Biolabs. Exonuclease III (2.5 or 10 units per reaction) was added to the PCR reactions as a mixture with Arrow™ Taq DNA polymerase (1 unit per reaction). Storage at −20°C of the mixture (i.e., Arrow™ Taq DNA polymerase plus exonuclease III) for several months did not alter the beneficial effect of the repair enzyme on Long PCR. QA-Agarose™ molecular biology grade was from Qbiogene.

A PCR protocol was set up to amplify either a 894-bp or a 3132-bp fragment on the human β-globin gene. PCR was performed in 0.5 ml thin-wall PCR tubes (Sarstedt, Nümbrecht, Germany) using a Crocodile III thermocycler equipped with a heat cover (Appligene, Illkirch, France). Amplification reactions were performed with Arrow™ Taq DNA polymerase (Qbiogene, Illkirch, France), which is a mixture of recombinant Taq DNA polymerase and recombinant Tfu DNA polymerase that possesses effective 3'-5' exonuclease activity. The final composition of the Arrow™ PCR buffer (provided as a 10 X mixture) is 50 mM KCl, 1.5 MgCl₂, 0.1% Triton X-100, 0.2 mg/ml BSA, 10 mM Tris-HCl pH 9.0. In some experiments, amplification was performed with the recombinant Taq DNA polymerase alone (Qbiogene).

2. PCR Conditions

Forward primer A 5'-TGGGCATAAAAGTCAGGGCAG-3' (nucleotides 62101 to 62121 on the human β-globin gene) and backward primer B 5'-ACATCCTCCAGGTTCATCA ATT-3' (nucleotides 65232 to 65211) amplified a 3132-bp fragment. Forward primer A and backward primer C 5'-GCAAATAAGCACATATATTCCAAA-3' (nucleotides 62994 to 62969) amplified a 894-bp fragment from the β-globin gene.
PCR reactions were performed in a volume of 50 µL, with 50 pmol of primers, total DNA (1 and 100 ng for amplification of the 894-bp and 3132-bp fragments, respectively), each dNTP (100 µM), magnesium chloride (1.5 mM; already present in the incubation buffer), and Arrow™ Taq DNA polymerase (1 unit). In some reactions, Taq DNA polymerase (1 unit) was used instead of Arrow™ Taq DNA polymerase. Exonuclease III (2.5 or 10 units) was added along with Arrow™ Taq DNA polymerase in some of the reactions.

1. For 10 reactions, prepare mix A and keep on ice:
   - Sterile distilled water 410 µl
   - 10 × Arrow™ PCR buffer 50 µl
   - dNTPs mix (5 mM each) 10 µl
   - Primer A (100 pmol/µl) 5 µl
   - Primer B or C (100 pmol/µl) 5 µl

2. Add 48 µl of mix A to a 0.5 ml thin-wall PCR tube kept in ice. Add 1 µl of DNA sample and then 1 unit of Arrow™ Taq DNA polymerase (supplemented or not with exonuclease III).

3. Place the tubes in the thermocycler and start the PCR program. To amplify the 894-bp fragment from the human β-globin gene, the thermocycler profile includes initial denaturation at 93°C for 5 min, followed by 35 cycles of 91°C for 1 min, 62°C for 1 min, 72°C for 1.5 min, and no final extension. To amplify the 3132-bp fragment from the β-globin gene, the PCR profile includes initial denaturation at 93°C for 5 min, followed by 25 cycles of 94°C for 30 sec, 62°C for 2 min, 72°C for 3 min, without final extension.

4. At the end of the PCR program, add 12.5 µl of a blue stop solution (50% glycerol; 0.1% bromophenol blue; 1% SDS; 50 mM EDTA; 50 mM Tris-HCl pH 7.5) into each tube.

3. **Analysis of PCR Products and Interpretation of Data**

   PCR products (12 µl) are electrophoresed on ethidium bromide–containing 1% QA-Agarose™ gels and photographed under UV transillumination.

   Figure 2.2 shows the beneficial effect of exonuclease III on the amplification of the 3132-bp fragment from the human β-globin gene with DNA samples that were damaged either intentionally (Figure 2.2A) or unintentionally (Figure 2.2B). With DNA samples heated at 99°C for 120 sec prior to the PCR experiments, addition of exonuclease III into the PCR medium increased the yield of the 3132-bp amplicon and also enhanced the specificity of amplification, with almost total disappearance of spurious products and smears (Figure 2.2A). This beneficial effect was higher with 10 than with 2.5 units of exonuclease III (Figure 2.2A). With a DNA sample that had been involuntarily thawed due to freezer failure, exonuclease III supplementation also increased the yield of the 3132-bp product and improved the specificity of the PCR reaction (Figure 2.2B).

   Figure 2.3 shows that exonuclease III also enhances the amplification of shorter amplicons from damaged DNA templates. With heat-damaged DNA samples, the addition of 2.5 and 10 units of exonuclease III again increased the yield of a 894-bp fragment from the human β-globin gene. As previously noted, the beneficial effect of exonuclease III was already evident with unheated DNA samples, suggesting that even carefully prepared DNA samples can contain DNA lesions blocking the progress of polymerases and limiting PCR amplifications. In general, Arrow™ Taq polymerase (Taq plus Tfu DNA polymerase) provided better amplification than Taq polymerase alone (probably due to the 3'-5' exonuclease activity of the Tfu polymerase). However, for the most damaged DNA sample (heated at 99°C for 120 sec), the 894-bp PCR product was almost absent with Taq alone or Taq and Tfu polymerases, although excellent amplification was obtained by adding exonuclease III (Figure 2.3).
C. Common Pitfalls, Troubleshooting Tips, and Specific Remarks

It should be noted that there are several experimental circumstances when exonuclease III has no beneficial effect or can even decrease amplification. First, when DNA templates are too severely damaged, exonuclease III cannot restore amplification.4

Second, as already discussed in relation to the results reported in Figure 2.1B, exonuclease III can decrease amplification of DNA fragments when both low amount of primers and an excess of exonuclease III are present in the PCR medium.4,5 Exonuclease III-mediated degradation of the primers can be overcome by increasing their amounts (when feasible), decreasing exonuclease III concentration, or preheating exonuclease III (99°C for 10 min) before the PCR experiment.4 Even though these modifications can restore amplification of the 316-bp mtDNA fragment,4 exonuclease III supplementation usually does not further increase the yield of small amplicons from damaged DNA samples. Thus, although exonuclease III is useful for Long PCR, it is not beneficial for the amplification of short DNA fragments. Qbiogene is currently performing experiments to determine the minimum amplicon length for which exonuclease III supplementation has reproducible beneficial effects.

Finally, it should be emphasized that the amount of exonuclease III necessary to give optimal results may depend on several factors, including the length of the amplicon, the nature of the DNA target (mtDNA, nuclear DNA, plasmids), and the DNA polymerases that are used in the reaction.

---

**FIGURE 2.2** Exonuclease III enhances amplification of a 3132-bp fragment of the human β-globin gene, with heat-damaged DNA templates. (A) Total DNA from one freshly drawn human blood sample was divided into different aliquots that were heated at 99°C for 0, 60, 90, or 120 sec prior to Long PCR amplification of a 3132-bp sequence of the human β-globin gene. Amplification was performed with the Arrow™ Taq DNA polymerase mix (Qbiogene) supplemented or not with 2.5 or 10 units of exonuclease III, as indicated. With the most damaged DNA sample (heated at 99°C for 120 sec) and in the absence of exonuclease III, the 3132-bp product was greatly decreased compared to that of undamaged templates. This product was restored by exonuclease III supplementation, which also enhanced the specificity of amplification, as shown by a marked decrease in spurious amplification products. (B) DNA was isolated from a human blood sample that had been thawed after freezer failure. The 3132-bp fragment was amplified as described above. Supplementation with 2.5 or 10 units of exonuclease III increased both the yield of the 3132-bp product and the specificity of amplification. M, Raoul™ marker (Qbiogene) which contains 22 DNA fragments, including a 2938-bp fragment migrating close to the 3132-bp PCR product. Shown are negative images of the original gel photograph.
IV. CONCLUSIONS

DNA lesions can occur in living organisms due to genotoxic chemicals, radiation, oxidative stress, and certain pathological conditions. DNA strand breaks, AP sites, and other blocking DNA lesions block or hamper DNA polymerases during PCR amplification. The larger the targeted sequence, the greater the probability that one or several blocking lesions will be present on each DNA molecule, thus selectively impairing the amplification of long DNA fragments.

This selective impairment can be used to detect blocking lesions on DNA after genotoxic treatments (Figure 2.1 to Figure 2.3). A 4-primer Long PCR methodology allows the concomitant amplification of small and large mtDNA fragments. This approach has the advantage of being rapid (results can be obtained in a few hours), sensitive (a few ng of total DNA can sometimes be enough), specific (provided by the use of primers to amplify a specific target sequence), and easy to quantify (with Poisson’s equation). Another advantage of the technique is its low running costs once experimenters have invested in a good quality thermocycler. Finally, although it cannot yet identify the nature of the blocking lesions, the use of selected DNA repair enzymes, including exonuclease III, might allow this question to be addressed in the future.

Besides genotoxic treatments and DNA-damaging conditions, DNA lesions can also occur postmortem, during tissue conservation, DNA extraction, or DNA storage. These secondary DNA alterations are major problems for forensic and archeological studies. They can affect also other studies when DNA is not optimally prepared or is accidentally damaged during storage but long DNA stretches need to be amplified.

FIGURE 2.3 Exonuclease III enhances amplification of a 894-bp fragment from the human β-globin gene with heat-damaged DNA samples. Total DNA from one freshly drawn human blood sample was divided into different aliquots that were heated at 99°C for 0, 60, 90, or 120 sec prior to Long PCR amplification of a 894-bp sequence of the human β-globin gene. Amplification was performed with the Arrow™ Taq DNA polymerase mix (Qbiogene) or with Taq polymerase alone (Qbiogene). These polymerases were supplemented or not with 2.5 or 10 units of exonuclease III, as indicated. Exonuclease III enhanced amplification of the 894-bp fragment with the heat-damaged DNA samples and the unheated DNA sample. With the DNA sample heated at 99°C for 120 sec, no amplification of the 894-bp PCR product was obtained with either Taq polymerase alone or the Arrow™ polymerase mix, but the fragment was retrieved with exonuclease III supplementation. Note that for this 894-bp PCR product, the enhancing effect of exonuclease III is reproducibly higher with 10 than with 2.5 units of exonuclease III.
Therefore, the discovery that exonuclease III (a major DNA repair enzyme of E. coli) enhances Long PCR amplification of DNA templates with DNA strand breaks and abasic sites (Figure 2.1 to Figure 2.3 and figures in Reference 4) is a major technological advance. Even though the mechanisms of this beneficial effect are still hypothetical, this discovery has two important applications. First, addition of exonuclease III to PCR reactions allows better amplification of Long PCR products from mildly damaged DNA samples, and allows amplification of more severely damaged DNA samples that would not be amplified without exonuclease III (Figure 2.1 and Figure 2.3). This can be critical when the PCR product must be used for further sequencing, cloning, or restriction analyses. In this regard, exonuclease III supplementation does not decrease the fidelity of Arrow™ Taq polymerase (P. Schmitt, personal communication), but further experiments are needed to extend this result to other DNA polymerases. Second, exonuclease III can be used to confirm the presence of abasic sites and/or strand breaks on a DNA target. Finally, since low amplification can be due to DNA damage or the presence of PCR inhibitors, exonuclease III supplementation and mixing experiments can reveal the cause of low amplification yield.

Qbiogene is developing a commercial kit containing blends of polymerases and exonuclease III optimized for the amplification of damaged DNA templates. Hopefully, these new kits will help scientists to retrieve DNA sequences from DNA samples that have been unintentionally damaged.

REFERENCES

11. Nakamura J. et al., 5'-Nicked apurinic/apyrimidinic sites are resistant to β-elimination by β-polymerase and are persistent in human cultured cells after oxidative stress, J. Biol. Chem., 275, 5323, 2000.
15. Lindahl, T. and Andersson, A., Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid, Biochemistry, 11, 3618, 1972.


3 Quantitative mRNA Analysis in Small Cell Samples by RT-PCR and Flow Cytometry

Veronika Stemme and Sten Stemme

CONTENTS

I. Introduction ................................................................................................................. ............ 23
II. Materials and Methods ....................................................................................................... .... 24
   A. Isolation and Culture of Endothelial Cells ................................................................. 24
   B. Endothelial Cell Sorting by Flow Cytometry ................................................................. 24
   C. Preparation of Cell Lysates for RT-PCR ................................................................. 25
   D. Counting of Cell Nuclei by Flow Cytometry: Calibration of the Flow Cytometer ...... 25
   E. RT-PCR, Primers, and Internal Standard ........................................................................ 26
III. Results and Discussion ..................................................................................................... ...... 26
Acknowledgments ............................................................................................................................ 27
References ........................................................................................................................................ 27

I. INTRODUCTION

The expression rate of key genes in specific cell populations in vivo, such as obtained from surgical biopsies, is of great interest. The polymerase chain reaction (PCR) technique in combination with a cDNA synthesis step using reverse transcriptase can be used to detect very low numbers of mRNA copies. Several methods also allow reliable quantitation of small amounts of mRNA. In order to measure rate of expression in cellular samples, the quantitative estimates should most appropriately be related to the number of cells in the original sample. Cell counting in small samples, however, constitutes a methodological problem. Modern flow cytometers are commonly used to measure cell surface proteins after antibody staining and can also sort cells according to specific cell surface markers. These features led us to test whether flow cytometry could be used in conjunction with quantitative RT-PCR for the estimation of mRNA copies per cell. Freshly isolated endothelial cells were sorted by flow cytometry and then lysed. Endothelial cells in primary culture were lysed directly in the microtiter plate wells. The number of cells in the lysate was determined by counting of nuclei after propidium iodide staining using flow cytometry. Plasminogen activator inhibitor-1 (PAI-1), an important regulator of the fibrinolytic system, is expressed by endothelial cells and was used here as an example target molecule. The number of PAI-1 mRNA copies per cell was determined by quantitative RT-PCR using point-mutated PAI-1 cRNA as an internal standard, although other quantitative RT-PCR methods could also be combined with the quantitative flow cytometry. The method allowed reliable and reproducible estimation of the number of target mRNA copies per cell from original cell samples containing less than 1000 cells. This method can be used for the quantitative determination of mRNA in specified cell populations from small tissue samples or cultured cells.
II. MATERIALS AND METHODS

A. ISOLATION AND CULTURE OF ENDOTHELIAL CELLS

Human umbilical vein endothelial cells (HUVEC) were isolated from umbilical cord veins by collagenase digestion as described earlier. Collagenase type I (Sigma, St Louis, MO) 200 U/μL in phosphate buffered saline (PBS, pH 7.2) with the addition of 5 mg/μL bovine serum albumin (RIA grade, Sigma), 1 mg/μL glucose and 0.2 mM CaCl₂ was used. The resulting cell suspension was either directly sorted by flow cytometry and analyzed by RT-PCR or seeded in 96-well microtiter-plates (Nunc, Roskilde, Denmark) precoated with 0.2% gelatin. The cells were cultured in RPMI 1640 culture medium (Life Technologies, Paisley, Scotland) with the addition of 20% fetal calf serum, 20 IU/μL heparin (Kabi Pharmacia, Uppsala, Sweden), and 20 μg/μL endothelial cell growth factor (Boehringer Mannheim Biochemica, Mannheim, Germany) at 37°C in 5% CO₂. HUVEC in primary culture, i.e., never subcultured, were used for quantification of PAI-1 mRNA.

B. ENDOTHELIAL CELL SORTING BY FLOW CYTOMETRY

The lectin *Ulex europaeus* I agglutinin (UEA I) recognizes carbohydrate antigens which are primarily expressed on endothelial cells and is commonly used as an endothelial cell marker. For

![Flow chart for the isolation of cells and processing of directly sorted cells (A) and after cells analyzed after primary cell culture (B). From Stemme, V. et al., Quantitative analysis of specific mRNA species in minute cell samples by RT-PCR and flow cytometry, *J. Imm. Methods*, 249, 223, 2001, with permission from Elsevier Science.]
flow cytometry, freshly isolated endothelial cells (EC) in umbilical vein collagenase digests were identified by staining with UEA I conjugated with fluorescein isothiocyanate (FITC) (Vector Laboratories, Burlingame, CA). To enhance separation between ECs and the erythrocyte population present in the cell suspension, the erythrocytes were labeled with a phycoerythrin-conjugated mouse monoclonal antibody against glycoporphin A (Dakopatts, Glostrup, Denmark). ECs were sorted with a FACSort flow cytometer (Becton Dickinson Immunocytometry Systems, San Jose, CA) into 50 µL tubes precoated with 4% bovine serum albumin (Sigma) in PBS. The sorted cells, typically in 50 to 150 µL PBS, were centrifuged at 400 g for 10 min. The pellets were resuspended in PBS, transferred to 1.5-µL tubes, and centrifuged at 400 g for 5 min. The supernatant was discarded using a 0.8 mm needle and the cells were lysed in lysis buffer as described below. Other cell types can be sorted using other cell-specific markers.

C. Preparation of Cell Lysates for RT-PCR

Two different protocols were used for preparation of lysates from freshly isolated EC sorted by FACS and from EC cultures respectively, as illustrated in Figure 3.1. In both protocols the cells were lysed in 100 µL of lysis buffer containing 0.05% NP-40, 1 U/µL RNase Inhibitor, 1 mM dithiothreitol, 140 mM NaCl, 1.5 mM MgCl₂, 10 mM Tris-HCl, pH 8.0. NP-40 was found to have an inhibitory effect on the efficiency of the RT-PCR reaction at 0.5% concentration in the lysis buffer. Titration experiments were therefore performed and 0.05% was found to be optimal for RT-PCR and counting of nuclei.

Freshly isolated endothelial cells were purified by flow cytometry cell sorting before lysis, while EC from primary cultures were washed with PBS and lysed directly in the microtiter plate wells by addition of lysis buffer. An aliquot of 10 µl was taken from each lysate for counting of cell nuclei by flow cytometry. As a nuclear stain, a concentrated stock solution of the DNA-binding fluorochrome propidium iodide was added to a final concentration of 0.5 µg/µL. The nuclei in the remaining samples were removed by centrifugation at 10,000 g for 5 min and the supernatants were stored at -20 °C until analyzed by RT-PCR.

D. Counting of Cell Nuclei by Flow Cytometry: Calibration of the Flow Cytometer

The flow cytometer was used to count cells in cell lysates by counting nuclei after lysis with detergent and nuclear staining with propidium iodide, as described above. The flow cytometer continuously draws sample from the sample tube without indication of volumes. In order to allow absolute counts using the analysis time and the sample flow to calculate cell concentration, and thus absolute numbers of cells in a known volume, the stability of the sample flow has to be documented. For this purpose, calibration beads (Cytek Development, Fremont, CA) with a known concentration were serially diluted from 20,000 beads/150 µL to 625 beads/150 µL, and analyzed at three different occasions. The samples were analyzed at the high sample flow setting for 60 sec using a timer function of the instrument and the actual sample flow was calculated using the formula:

\[
\frac{\text{counted beads in 1 min}}{\text{bead concentration (beads / µL)}} = \text{sample flow (µL / min)}
\]

Both between and within assay coefficients of variation were below 10% and the calculated sample flow corresponded well to the approximate flow specified by the manufacturer (60 µL/min). The results indicate that the flow cytometer can be used for determination of cell concentration using the sampling time and the fixed sample flow without need for calibration in each experiment.
E. RT-PCR, Primers, and Internal Standard

A 1.1 kb Pst I-fragment of PAI-1 cDNA in a pGEM-3 vector (Promega, Madison, WI) was used for the synthesis of an internal standard for quantitative PCR, according to Ikonen et al. A single nucleotide was mutated (nt 182, C→G) using PCR and the plasmid with the mutated fragment was then used for in vitro RNA synthesis. Known amounts of synthetic RNA standard were reverse-transcribed and amplified together with sample RNA using a biotinylated 5’ primer, 5’- CGC AAG GCA CCT CTG AGA ACT -3’ (PAI-101; nt 50-70) and a 3’ primer, 5’- CGT GCT CCG GAA CAG CCT GAA -3’ (PAI-102; nt 484-504). The RT and PCR were performed in the same tube in 100 μL of 20 mM Tris-HCl, pH 8.4, 2 mM MgCl₂, 60 mM KCl using 0.2 U/μL RNase Inhibitor (Boehringer Mannheim), 25 U StrataScript Reverse Transcriptase (Stratagene, La Jolla, CA), 2.5 U Taq DNA Polymerase (Boehringer Mannheim), dNTPs at 0.2 μM each, 10 pmol of the biotinylated PAI-101 primer, and 25 pmol of the PAI-102 primer. The reaction mixtures were incubated at room temperature for 10 min and the RT was performed at 42°C for 1 h followed by denaturing at 94°C for 5 min. Thirty PCR cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 20 sec + 1 sec/cycle were performed. The biotinylated PCR product was immobilized on a streptavidin-coated solid support (microtiterplates by Wallac, Finland) and denatured. The ratio between the two amplified sequences was determined by separate “minisequencing” reactions, in which a detection step primer, 5’-AAC ACC CTC ACC CCG AAG TCT -3’ (PAI-108; nt 183-203), annealing immediately adjacent to the site of the variable nucleotide, was elongated by a single 3H-labeled dNTP complementary to the mutated or wild-type nucleotide at the variable site. The amount of incorporated labeled nucleotide was measured in a MicroBeta Liquid Scintillator Counter (Wallac, Finland). The ratio between the two incorporated labeled nucleotides corresponded to the ratio between the wild-type and the mutated standard RNA in the original RNA sample.

To estimate the precision of the RT-PCR analysis, calibration curves were prepared using a series of samples with a fixed amount of synthetic wild-type RNA together with varying amounts of synthetic mutated RNA. Repeated analysis of the same RNA samples over several days resulted in calibration curves which were highly reproducible. The calibration curves deviated from the expected ideal curve at wild type internal standard ratios below 0.1 and above 5. Therefore, ratios between 0.80 and 1.20 were used for quantifications of mRNA. Inter-assay coefficients of variation were below 11% within this range. In order to obtain ratios within this interval, three dilutions of internal standard were run with each sample in the RT-PCR. Complete calibration curves or selected control samples were run in each RT-PCR assay to control for inter-assay variability.

III. RESULTS AND DISCUSSION

The flow chart for the quantitation of PAI-1 mRNA in freshly isolated or primary culture umbilical cord endothelial cells is shown in Figure 3.1. The results from three separate experiments are shown in Table 3.1. PAI-1 mRNA expression is clearly detectable in all cases. Quantitative RT-PCR allows quantification of mRNA from samples corresponding to a few hundred cells. The result is often related to the amount of total RNA added in the initial RT-PCR reaction or to the amount of a “house-keeping gene.” These approaches, however, have important disadvantages. First, the number of mRNA copies per cell is in most cases of greater biological relevance. The total RNA content and also the expression of “house-keeping genes” vary widely between different cell types, between tissues and cultured cells, with the differentiation state and with the metabolic state induced by stimulatory agents. Second, this relationship to the amount of total RNA sets a limit as to how small cell samples can be quantified since the RNA isolation must result in amounts of RNA large enough to be quantified by conventional methods, most commonly spectrophotometry. Attempts to overcome this by preparing RNA from a known amount of cells are hampered by the unpredictable yield of most RNA isolation methods, particularly from small cell samples. RT-PCR directly from cell lysates and tissue samples without previous RNA isolation is possible. However, the problem
Quantitative mRNA Analysis in Small Cell Samples by RT-PCR and Flow Cytometry

TABLE 3.1
Expression of PAI-1 mRNA in Freshly Isolated Endothelial Cells from Human Umbilical Vein

<table>
<thead>
<tr>
<th>Experiment</th>
<th>PAI-1 mRNA (Copies/Cell)</th>
<th>Cell Number (Nuclei)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>390</td>
<td>2690</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>850</td>
<td>400</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>764</td>
<td>500</td>
</tr>
</tbody>
</table>


of estimating the amount of total RNA or cell number in the original sample remains (see also Chapter 22 by Stephen A. Bustin).

Quantitative flow cytometry makes it possible to reliably quantify the number of cells in cell lysates. Used together with a quantitative RT-PCR, this method allows estimates of the number of mRNA copies per cell in minute cell samples. For the quantitative RT-PCR we employed a method based on the use of an internal standard differing from the target molecule by a single nucleotide and quantification in a final minisequencing reaction, described by Ikonen et al.10 This method was applied on cell lysates where the number of cells could be determined by counting the nuclei by flow cytometry. The procedure circumvents the necessity for previous isolation of RNA and at the same time allows quantification of cell numbers in the original sample. It allows estimation of numbers of mRNA copies per cell in original samples consisting of less than 1000 cells. Combined with flow cytometry cell sorting this method opens up the possibility to quantitatively determine the in vivo production of various mRNA species in selected cell populations from minute clinical or experimental samples. Furthermore, the sensitivity of the method makes it possible to quantify specific mRNAs in small-scale experiments on short-term primary cell cultures. Cells may change their phenotype during culture procedures,15,16 and primary cultures presumably constitute a better model for cells in vivo.

There may be a concern that the cell sorting procedure could affect the mRNA levels, although in our system we found comparable levels of mRNA in sorted and unsorted cells. Different methods for fixation of cells could be tried to circumvent this potential problem.17,18

The method presented here requires a flow cytometer in addition to common RT-PCR equipment. The minisequencing protocol requires a liquid scintillator counter for microplates and an internal standard.

ACKNOWLEDGMENTS

A 2.7 kb fragment of PAI-1 cDNA9 in a pUC 8 plasmid was a kind gift from Dr. V. W. M. van Hinsbergh (Gaubius Laboratory, IVVO-TNO, Leiden, The Netherlands).

REFERENCES

I. INTRODUCTION

PCR has been used to detect nuclear DNA or reverse transcribed expressed RNA sequences from extracts of various tissues. For many purposes, greater specificity and sensitivity can be achieved if microdissection is used to collect groups of similar normal or pathological cells from tissue sections for nucleic acid extraction. An example of where single cell PCR is required is for the study of somatic
mutations such as in the antibody chains of B lymphocytes, because the sequences of rearranged DNA for antibody heavy chains of each B cell are unique. Direct sequencing of amplified DNA from a single cell minimizes errors from in vitro recombination during PCR. This would be a particularly serious problem for studies of antibody heavy and light chain sequences in rabbit B cells where blocks of changes occur that are introduced by a gene-conversion-like mechanism. We started using hydraulic micromanipulation (HM) for single cell collection as described by other laboratories (reviewed in Reference 5). Most of the results we published used cells collected by HM. Since such manual microdissection is extremely difficult and tedious to perform, we turned to laser capture microdissection (LCM). The methods developed to stain and analyze cells for direct comparisons of different microdissection methods are given in detail below. Results of direct comparisons of yields of PCR products and sequences indicative of recovery of a single cell collected by HM or LCM from the same stained sections were published. We concluded that the reproducibility, better yields of single cell sequences, greater efficiency, and shorter times needed for both operator training and actual cell collection made LCM a valuable method for single cell collection. In addition to tests of the infrared laser-based LCM instrument manufactured by Arcturus Engineering, we also tested the Leica UV laser-based LMD instrument, doing a similar type of comparison. The ease of use of the noncontact Leica LMD system is very attractive. However, the yields of PCR products and single-cell sequences were unacceptably low. On the other hand, isolation of RNA from groups of cells such as the contents of a germinal center in serial sections from spleen was simple and successful by the Leica LMD method (Yang G. et al., unpublished results).

II. EXAMPLES

A. INSTRUMENTATION AND SUPPLIERS

1. Laser-Based Microdissection Systems
   a. PixCell®Ie Laser Capture Microdissection System and Image Archiving Workstation, (LCM) [Arcturus Engineering, Mountain View, CA (www.arctur.com)]

2. Hydraulic Micromanipulation (all instruments from Narishige, Tokyo, Japan)
   Grinder BG-40
   Microforge MF-900
   Needle Puller PC-10
   Micromanipulator MMN-1

3. Cryostat Microtome
   Tissue-Tek™ cryostat microtome (Sakura Finetek USA Inc., Torrance, CA).

B. REAGENTS AND SUPPLIERS

1. Precleaned microscope slides (Fisher Scientific, Pittsburgh, PA)
2. Polyethylene naphthalene (PEN)-foil coated microscope slides (Matsunami Glass, Kishiwada City, Japan)
3. Placental RNase inhibitor (Calbiochem, La Jolla, CA)
C. EXPERIMENTAL PROTOCOL

1. Tissue Collection and Sectioning

Human and rabbit appendix or rabbit spleen tissues were cut into 2 to 3 mm fragments. They were embedded in optimal cutting temperature (OCT) compound and stored at –70°C. Sections of 8 mm were cut at –20°C using Tissue-Tek™ cryostat microtome. HM and LCM sections were mounted on precleaned microscope slides. LMD sections were on precleaned UV irradiated PEN-foil coated microscope slides. Sections for DNA extraction were immediately stored at –70°C while those for RNA extraction were first vacuum dried for 10 min before storage at –70°C.

2. Staining for DNA Extraction

Tissue sections were (1) removed from the –70°C freezer and slides air dried for 30 min to achieve optimal adherence; (2) acetone-fixed for 10 min; (3) rinsed with Cadenza Buffer; (4) blocked with 100 µl of 5% goat serum for 10 min to minimize nonspecific antibody binding; (5) incubated with 100 µl of primary antibody for 30 min [e.g., mouse antihuman CD79a to identify human or rabbit B lymphocytes (Pharminen, San Diego, CA)]; (6) rinsed in Cadenza Buffer; (7) incubated with 100 µl of secondary antibody [e.g., goat antimouse IgG-biotin (Southern Biotechnology, Birmingham, AL)] for 20 min; (8) incubated for 20 min with Vector ABC peroxidase and 10 min with DAB-Ni substrate to visualize stained B cells; and (9) counterstained with hematoxylin for better contrast.

3. Staining for RNA Extraction

Tissue sections were (1) removed from the –70°C freezer and vacuum dried for 10 min to minimize endogenous RNase activity; (2) acetone-fixed for 10 min; (3) rinsed with Cadenza Buffer prepared with DEPC water; (4) incubated with antigen coupled to alkaline phosphatase (AP) (e.g., DNP-AP antigen1–3) diluted in 100 µl of 5% goat serum containing 0.2U/ml placental RNAse inhibitor for 20 min; (5) rinsed with Cadenza Buffer prepared with DEPC water; (6) incubated for 15 min with Vector ABC-AP substrate kit II and treated with placental RNAse inhibitor (Vector Lab. Inc.) to visualize stained antigen-specific B cells; and (7) rinsed with DEPC water.

4. Dehydration

Stained tissues were treated with 75% ethyl alcohol for 30 sec, 95% ethyl alcohol for 30 sec, 100% ethyl alcohol, three changes for 2 min each, cleared in xylene, three changes for 2 min each, and air dried at room temperature for about 30 min.

5. Hydraulic Micromanipulation

The hydraulic micromanipulator (Narishige, Tokyo, Japan) requires micropipettes (ranging 1 to 10 µm in diameter) for the collection of single cells from tissue sections. Using this instrument, individual B lymphocytes from human or rabbit appendix germinal centers and rabbit splenic germinal centers were collected. Each of the targeted cells was individually removed from the tissue section with the micropipette, transferred to a 0.2 ml microfuge tube containing 5 µl of...
alkaline lysing solution (200 mM KOH/50 mM DTT), and incubated for 10 min at 65°C to extract the DNA.

6. Laser Capture Microdissection

Our previously published methods and results used the earlier PixCell® II instrument and a “noncontact” CapSure HS™ disposable cap with rails that were 4 μm high. These rail caps are no longer supplied by the manufacturer (Arcturus Engineering). Although we have much less experience with it, the current PixCell® Ile LCM system (Arcturus Engineering, Mountain View, CA) has now also been used to perform single cell LCM, image acquisition, PCR, and sequencing of rearranged heavy and light chain sequences from single B cells. The PixCell® Ile uses 12 mm “noncontact” CapSure HS™ disposable cap designed to reliably separate its nonactivated transfer polymer surface from the tissue surface within a central 3 mm diameter zone. Four large arcs of a ring 12 μm high hold the polymer surface above the tissue, thereby preventing contact and inadvertent cell pickup within the central zone. A black circle ~2.5 mm in diameter was drawn inside the 12 μm high arcs to mark the microdissection zone (Figure 4.1). For single cell LCM using our stained tissue sections, the PixCell® Ile, and 12 μm CapSure HS™ rail caps, the laser settings were in the following range: 60–65 mW, and duration 600–650 μsec. These settings were sufficient to focally melt the polymer film above the targeted cell within the black circle. The polymer then expanded across the 12-μm spacing to bond and capture the desired cell. Gentle lifting up of the cap isolated the targeted cell with nonspecific cells adhering only onto the region of rail contact. The cap was then placed face up in the CapSure™ HS alignment tray and the ExtracSure™ sleeve (Arcturus Engineering) was affixed firmly. This sealed the rails and excluded any material from the tissue section that may have adhered to the rails during contact, exposing only the targeted cell. The CapSure™ HS and ExtracSure™ assembly containing the captured cell in the exposed black circular area was used to cap a 0.5 ml microfuge tube during incubation in 5 μl of alkaline lysing solution.

**FIGURE 4.1** Diagram of the CapSure HS™ Cap showing a captured cell, the black circle, and the 12 μm rails. Shown above the cap are the ExtracSure™ device and the microfuge tube.
at 65°C for 15 min to extract the DNA. The DNA extract was then collected in the microfuge tube by centrifugation at 4000 x g for 2 min.

7. Leica AS LMD Laser Microdissection

PEN-foil slides were placed face down on the microscope stage and stained tissue was observed on a computer screen via a video camera. Visualized cells were targeted using 63 X for a single cell and 20 X objectives for cell clusters. An identified B cell or cell cluster was circled using the computer mouse, and a click at the “cut” command activated the UV laser to cut a trace around the circular path. The ablation forces, reactive forces, and generated charge effects in combination with gravitational force deposited the target cell or cluster into the cap of a 0.2 ml microfuge tube centered beneath the tissue section. The microfuge tube was capped and centrifuged for 2 min at 4000 x g to move the sample from the cap into the tube. DNA was extracted from samples by incubating in 5 μl of alkaline lysing solution at 65°C for 10 min. RNA was extracted from samples with the PicoPure™ RNA isolation kit (Arcturus Engineering, Inc.) following the manufacturer’s protocol.

8. Single-Cell PCR and DNA Sequencing

The DNA extract of the alkaline lysed cell was neutralized with 5 μL of neutralizing solution (900 mM Tris-HCl, pH 8.3, 300 mM KCl, 200 mM HCl). Neutralized samples were used to run nested, touchdown PCR amplification reactions.1–4,6

D. Interpretation of Data and Results

Table 4.1 summarizes our experience to date using the methods of microdissection described above and in1–4,6,7 to collect single B cells and to obtain PCR products of rearranged heavy and/or light chains of immunoglobulins that were indicative of collection of a single B cell. The efficiencies listed are for obtaining sequence indicative of collection of a single cell. Because each sequence is unique, when two or more cells were collected and forward and reverse strands of DNA were sequenced, a mixed sequence was obtained that did not assemble.

| TABLE 4.1 | Summary of Features of Different Methods for Microdissection |
|---|---|---|
| | HM | LCM | LMD |
| System specification | Manual manipulation using micropipettes | IR laser | UV laser |
| Ease of use | Very tedious | Less tedious | Very easy to use |
| | Time consuming | Requires one cap per cell | Automated system |
| | Nonreproducible | Requires elaborate extraction system | Samples collected in sets of four |
| Performance varies from day to day | | | |
| Requires prolonged training | | | |
| Visualizing captured cell | Not possible | Very easy | Not possible with the available 10 X objective |
| DNA damage | None | Probably none; (thermal possible) | Both thermal and photochemical possible |
| Efficiency | 55% | 75% | 18% |
Allow sections to air dry completely before microdissecting with LCM or LMD using the techniques described here. Partially hydrated sections will not allow the polymer to bond properly to the cell, thus hindering cell capture. In the case of LMD, the PEN-foil may stick to the glass slide, preventing the cut-out cell from dropping into the microfuge tube cap. Sections for RNA extraction should be vacuum dried immediately after sectioning and before immunohistochemistry when removed from −70°C to minimize RNA degradation. Sterile technique should be observed from microdissection to the end of the PCR amplifications. Avoid using more than 2.5 μl of the first round product to run the second round PCR reaction, as greater volumes will introduce considerable variability in the concentrations of the second round reagents. When using the PixCell IIe for single cell LCM, it is better to calibrate the laser using higher power settings and durations and then slowly step down to the lowest possible settings that give the desired result. For example, start with 70 mW power and 700 μsec duration and work your way down to achieve optimum power setting (60–65 mW and duration of 600–650 μsec under our conditions). Another technique variation for highly specific capture is to further reduce this single-pulse power to 70–80% of this threshold and deliver a series of pulses until the polymer reaches and bonds the target cell. For better visualization of the captured cell after separation of the cap from the tissue, divide the power setting used to microdissect the cell by two, multiply the duration by two, and use this setting to flatten the cone around the captured cell (Figure 4.2). Extracting samples from caps observed under the microscope to contain only the single captured cell greatly improves the likelihood of obtaining a PCR product and single-cell sequence.

III. DISCUSSION

A. BUDGET, EXPERTISE, AND EFFORT REQUIRED

As summarized in Table 4.1, hydraulic micromanipulation is the most difficult to master and most tedious to perform. On the other hand, the equipment is not as expensive as those that utilize laser technology. Because of their high cost, laser microdissection instruments are generally in core facilities [see e.g., the description of the Laboratory of Pathology, NCI Core facility at the National Institutes of Health, Bethesda, MD (http://home.ncifcrf.gov/ccr/lop/Clinical/lcm/default.asp)].

B. OTHER USEFUL APPLICATIONS OF THE METHOD

Microdissection for analyses of pathological and normal areas of patient tissues has broad applications. For example, after we successfully used the methods described above for analyses of immunoglobulin heavy and light chain sequences from single B cells, we also collaborated in studies of the somatically acquired Kit activating mutation D816V found in patients with
mastocytosis. A nested PCR followed by restriction enzyme digestion of the second round PCR product distinguished wild type from somatically mutated gene sequence. We first used the PCR method with cultured cells to demonstrate that the mutation enhances stem cell factor-dependent chemotaxis. Next we used the same PCR method to examine DNA extracts of microdissected mast cells, B and T lymphocytes (400 to 500 cells) from lesional and non-lesional areas of bone marrow tissue sections from mastocytosis patients. These were successfully amplified and analyzed for presence of the Kit somatic mutation (Taylor, M.L. et.al., manuscript in preparation).

C. Potential Future Developments

Laser capture microdissection is a new technology that has been rapidly commercialized into generally useful formats. Several commercial improvements could make comprehensive mapping of individual cell variations among cells (e.g., B lymphocytes within a germinal center or specific cells within an abnormal cytology or biopsy specimen) easier, quicker, and more reliable. For example, we have developed LCM prototypes that can capture any desired set of individual cells in such specimens onto a flexible noncontact LCM tape at separate independently specified positions appropriate for individual analyses. These locations can be mated with parallel extraction and PCR analysis microfluidics channels to provide for rapid mapping of macromolecular differences. Incorporating ligands with molecular specificity within the capturing polymer matrix may also provide for highly efficient surface affinity purification and amplification and eventually direct optical readout of individual results.

Acknowledgments

We thank Devinder Sehgal, Enrico Schiaffella, Arash Malekafzali, and Joseph F. Dasso for their contributions to development of the methods described. Barbara A. Newman and Michael Mage made helpful suggestions about the manuscript. Shirley Starnes provided expert editorial assistance.

References


5 Pre-PCR Processing Strategies

Peter Rådström, Maria Lövenklev, Petra Wolflis, Charlotte Löfström, and Rickard Knutsson

CONTENTS

I. Introduction................................................................................................................. 37
II. Sample Preparation......................................................................................................... 38
   A. Biochemical Methods................................................................................................. 38
   B. Enrichment Methods ................................................................................................. 39
   C. Immunological Methods ........................................................................................... 39
   D. Physical Methods ....................................................................................................... 39
III. DNA Polymerases.......................................................................................................... 40
IV. PCR Facilitators............................................................................................................ 40
V. Concluding Remarks...................................................................................................... 43
Acknowledgments ................................................................................................................ 43
References ............................................................................................................................. 44

I. INTRODUCTION

PCR is particularly valuable for monitoring gene expression, quantifying food-borne pathogens, testing viral load, and also for clinical diagnosis. However, although it can be extremely effective with pure solutions of nucleic acids, its usefulness is limited, in part, by the presence of inhibitory substances. Originating from the samples or from DNA extraction protocols, these can reduce or even block amplification. Although many biological samples have been reported to inhibit PCR amplification, the biochemical and physical mechanisms and identities of many inhibitors remain unclear.

The thermostable DNA polymerase is probably the most important target of PCR-inhibitory substances.\(^1\) It is well known that the polymerization activity of DNA polymerases varies with temperature, ionic strength, pH, buffer ions, sulfydryl content, and other chemical agents.\(^2\) Another reason for the failure of PCR amplification is the status of the nucleic acids. For instance, any large or small molecule that binds to ssDNA or dsDNA may alter the melting temperature \(T_m\).\(^3\) This type of PCR inhibition can be caused by the presence of human immunoglobulin G, making the target nucleic acids unavailable to PCR.\(^4\) Other causes of PCR inhibition are chelation of \(\text{Mg}^{2+}\) or interference by ions such as \(\text{Ca}^{2+}\) which compete with \(\text{Mg}^{2+}\) in binding to the DNA polymerase.\(^5\) Heme and lactoferrin inhibit PCR through their ability to release iron ions.\(^6\) Thus, the PCR inhibitors may act through one or more of the following mechanisms:\(^7\) (1) inactivation of the thermostable DNA polymerase, (2) degradation or capture of the nucleic acids, (3) interference with the cell lysis step, and (4) interference one of the reaction components.

With the development of real-time PCR, the risk of PCR inhibition has increased. The addition of fluorogenic substances, such as amplicon probes or fluorescent dyes, has made the systems more complex, and the targets for PCR inhibition have increased in number. It is therefore not surprising that nearly all reports on the use of real-time PCR involve DNA extraction since this procedure is
one of the most thorough in the removal of PCR inhibitors. In addition to inhibition of the amplification reaction, factors interfering with the detection method have to be considered. Autofluorescence\textsuperscript{9} is a fluorescent signal generated by something other than the target DNA. For example, when minor groove binding dyes such as SYBR Green I (Roche Diagnostics, Mannheim, Germany) are used, the presence of high amounts of nontarget DNA may result in such a high unspecific signal that it dominates over the signal of the specific product. Recently, an entirely new group of real-time PCR interfering factors has been identified. These are substances that have a background fluorescence of their own, such as MOPS, Tris, and Triton-X (unpublished data). Substances or complexes with a high absorbance can also cause PCR interference by scattering the excitation light; known examples of this are blood- or charcoal-based enrichment media.

Some PCR vessels have been adapted to increase the speed of heat transfer during cycling (i.e., capillaries used for the LightCycler\textsuperscript{TM} [Roche Molecular Biochemicals] or Smart Cycler\textsuperscript{®} reaction tubes used for the Smart Cycler\textsuperscript{®} [Cepheid, Sunnyvale, CA]). The higher surface-to-volume ratio of these adapted tubes might lead to increased binding or adherence of inhibitors to the walls, thus reducing their effect. However, larger particles from complex biological samples, for example denatured blood proteins or meat particles from homogenated samples, might also negatively affect the flow of PCR components through the narrow PCR tubes.

II. SAMPLE PREPARATION

The objectives of a sample preparation step are: (1) to reduce the size of the heterogeneous bulk sample to a small homogeneous PCR sample in order to ensure negligible variation between repeated sampling, (2) to concentrate the target organism to a concentration within the practical operating range of the PCR assay, and (3) to remove or neutralize substances that may interfere with the PCR amplification. These objectives influence the choice of sample preparation method. Many sample preparation methods are laborious, expensive, and time consuming, or do not provide the desired template quality. Since sample preparation is a crucial step in PCR, a large variety of methods has been developed, which affects the PCR analysis differently in terms of amplification efficiency, sensitivity, and specificity. Sample preparation methods used for PCR can be divided into four major categories: (1) biochemical, (2) enrichment, (3) immunological, and (4) physical. The choice of sample preparation method should be made bearing in mind the aim of the PCR analysis.

A. BIOCHEMICAL METHODS

The biochemical methods include extraction of RNA or DNA with organic solvents. The advantage of these methods is that a homogeneous sample of high quality is obtained for reverse transcription and PCR amplification. Most PCR inhibitors are removed since the template is usually purified and stored in appropriate buffers such as diethyl pyrocarbonate (DEPC)-treated water or tris ethylene diamine tetra-acetic acid (TE) buffer. The choice of biochemical method is determined by the aim of the PCR analysis and the sample being analyzed. Several different applications have been described in the literature, such as the detection of the West Nile virus in mosquitoes,\textsuperscript{10} Legionella DNA in respiratory samples,\textsuperscript{11} and Bordetella pertussis in clinical tissues,\textsuperscript{12} the quantification of enterovirus RNA in sludge samples,\textsuperscript{13} Geobacter spp. in sediments,\textsuperscript{9} glutamate receptor gene expression in vestibular brainstem and cerebellum,\textsuperscript{14} and human cytomegalovirus DNA in urine samples.\textsuperscript{15} Drawbacks with many RNA extraction methods are the poor quality of total RNA preparations and the degradation of mRNA during cell lysis. Total RNA methods based on mechanical disruption of the cells using, for example, glass beads in a bead mill (Mini Bead Beater; Biospec Products, Bartlesville, OK) or a homogenizer is often preferable to enzymatic digestion. Furthermore, many of the biochemical methods are laborious, costly, and may suffer from batch-to-batch variation with respect to the purity and concentration of the template after extraction. To improve the PCR detection after the DNA has been extracted, some additional sample pretreatments can be
Pre-PCR Processing Strategies

performed, such as magnetic capture hybridization (MCH)\(^\text{16}\) and silica binding of DNA\(^\text{17}\) (a detailed protocol for the latter can be found in Chapter 1 by Anne C. Stone on extraction and amplification of ancient DNA).

B. Enrichment Methods

Enrichment PCR involves cultivation of the target microorganism in a liquid medium prior to PCR. Enrichment PCR results in a dynamic detection range during cultivation, i.e., a time span for positive PCR detection.\(^\text{18}\) The aim of the enrichment culture is to provide detectable concentrations of viable target cells prior to PCR.\(^\text{19}\) Furthermore, enrichment cultures prior to PCR analysis serve additional purposes, including the dilution of PCR-inhibitory substances present in the sample matrix, dilution of dead target cells, and last but not least, the possibility of isolating the target cells for complementary biochemical tests. Various enrichment PCR procedures have recently been used for the detection of food-borne bacteria such as the Shiga toxin-producing \textit{Escherichia coli},\(^\text{20}\) \textit{Campylobacter} spp.,\(^\text{21}\) \textit{Clostridium botulinum},\(^\text{22}\) \textit{Salmonella} Mbandaka,\(^\text{23}\) \textit{Listeria monocytogenes},\(^\text{24}\) and \textit{Yersinia enterocolitica}.\(^\text{25}\) Enrichment culture prior to real-time PCR technology has also been developed for \textit{Salmonella} detection in poultry\(^\text{26}\) and the detection of root-mat-inducing \textit{Agrobacterium} in plant samples.\(^\text{27}\) Selective and nonselective agar and enrichment media have been used, and the specificity will depend on the medium characteristics. However, most media contain components that inhibit or interfere with PCR.\(^\text{28}\) Recently, a PCR-compatible enrichment medium for \textit{Y. enterocolitica} was developed, which removes the necessity for sample pretreatment.\(^\text{18}\) This allowed simplified, integrated liquid handling, a closed-tube, real-time PCR analysis procedure, and the detection of small numbers of unlysed \textit{Y. enterocolitica} in the presence of high concentrations of background flora.

C. Immunological Methods

These sample preparation methods are often based on the use of magnetic beads coated with antibodies.\(^\text{29}\) Because they are relatively easy to automate, their use as a quantitative sample preparation step prior to quantitative real-time PCR has become an important issue;\(^\text{10}\) i.e., \textit{Escherichia coli} O157:H7 has been separated from poultry carcass rinse using automated immunomagnetic separation (IMS) prior to microarray detection.\(^\text{31}\) Recently, immunomagnetic isolation and reverse transcription (immunobead RT) PCR have been developed, which have significant advantages over the previous RT-PCR methods.\(^\text{32,33}\) However, the specificity of these protocols depends on the specificity of the antibodies as well as the PCR. Furthermore, complex matrices, such as feces, can interfere with the interaction between antigen and antibodies.\(^\text{34}\) After immunocapturing, the sample requires lysis or washing prior to PCR.\(^\text{35}\) Immunological methods improve the concentration of the target organism, but the homogeneity of the PCR sample differs depending on the processing following capture. Usually, the template is of appropriate quality after these treatments. Since the specificity relies, to a certain degree, on the antibodies, false negative results can be obtained. Other disadvantages of these methods are that they are expensive, rather laborious, and time consuming.

D. Physical Methods

Aqueous two-phase systems,\(^\text{36}\) buoyant density centrifugation,\(^\text{37}\) centrifugation,\(^\text{38}\) dilution,\(^\text{39}\) and filtration\(^\text{40}\) are examples of methods that have been used as physical sample preparation methods prior to PCR. These methods are dependent on the physical properties of the target cell, for example, cell density and size. Density media, such as BactXtractor\(^\text{41}\) and Percoll,\(^\text{42}\) concentrate the target organism and remove PCR-inhibitory substances that differ in density. A combined filtration-enrichment culture followed by PCR has been developed for rapid detection of \textit{Campylobacter jejuni} and \textit{C. coli} in human feces samples where only bacteria that pass through the membrane can multiply in the enrichment medium.\(^\text{43}\)
Knutsson et al. have evaluated different physical sample preparation methods namely aqueous two-phase systems, buoyant density centrifugation, crude centrifugation, and dilution. In aqueous two-phase systems, PCR inhibitors and target cells were gently partitioned between two immiscible phases. For example, an aqueous two-phase system based on polyethylene glycol (PEG) 4000 and dextran 40 was used to detect *Yersinia enterocolitica* in pork and enrichment medium. The method is easy to perform, but phase separation may take 30 to 60 min. Buoyant density centrifugation has been shown to be a promising method if rapid detection is important. This treatment isolates whole cells that can be directly used as a PCR sample. The homogeneity of the sample differs depending on the composition of the biological matrix. Sample matrix components with the same density are not removed and may inhibit DNA amplification. However, the advantages of density centrifugation are its user-friendliness and a concentration of the target organism, increasing the sensitivity of detection.

### III. DNA POLYMERASES

A key component of PCR is the thermostable DNA polymerase, and any factor that compromises its enzymatic activity will result in decreased formation of specific PCR products. The polymerase can be degraded or denatured, or the enzymatic activity can be reduced by a wide variety of compounds present in biological samples. A number of DNA polymerases are commercially available, such as *rTth* and *Tth* isolated from *Thermus thermophilus* and DyNzyme from *T. brockianus*. AmpliTaq Gold and Platinum Taq are *T. aquaticus* preparations that are activated by the first DNA denaturation step (“built-in” hotstart). The choice of DNA polymerase is determined by several factors related to the application and has been shown to influence the performance of several PCR-based applications such as restriction fragment length polymorphism (RFLP), short tandem repeat (STR), and randomly amplified polymorphic DNA (RAPD), as well as multiplex PCR assays, differential display RT-PCR, autosticky PCR, and real-time PCR.

The various DNA polymerases differ in many features that are important in PCR such as fidelity, the ability to incorporate modified bases, and termination of primer extension. Furthermore, recent research has indicated that different polymerases have different susceptibilities to PCR inhibitors. Some DNA polymerases, i.e., *Taq*, are susceptible to inhibition by biological samples, including clinical, environmental, and food samples, such as blood, cheese, feces, and meat, as well as various ions. The inhibition of PCR by components of biological samples can be reduced or eliminated by choosing an appropriate thermostable DNA polymerase without the need for extensive sample processing prior to PCR. Several studies have evaluated the usefulness and characteristics of different DNA polymerases with respect to various PCR samples (Table 5.1). PCR detection was significantly improved by replacement of *Taq* DNA polymerase by *Tth* DNA polymerase for assays detecting *Staphylococcus aureus* in bovine milk, the poultry pathogen *Mycoplasma iowae*, *Helicobacter hepaticus* in mice feces, as well as influenza A in clinical specimens. Furthermore, relief of inhibition from pig feces and growth media has been observed using the DNA polymerase *rTth*. The use of hotstart enzymes, such as AmpliTaq Gold, has been shown to increase the product yield for highly degraded DNA from paraffin-embedded tissue and in several studies to relieve inhibition from blood. These examples clearly demonstrate that the use of alternative DNA polymerases can improve the performance of PCR and relieve the inhibition from biological samples.

### IV. PCR FACILITATORS

Certain compounds, called amplification enhancers or amplification facilitators, can improve the performance of PCR when added to the basic PCR master mix. They can be divided into five
# TABLE 5.1
Examples of Pre-PCR Processing Strategies

<table>
<thead>
<tr>
<th>Sample</th>
<th>Recognized Inhibitor</th>
<th>Pre-PCR Processing</th>
<th>DNA Polymerase</th>
<th>Facilitator</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Food</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alfalfa seed</td>
<td>—</td>
<td>Enrichment and immunomagnatic separation</td>
<td>Taq (Roche Diagnostics)</td>
<td>Gelatin</td>
<td>23,72</td>
</tr>
<tr>
<td>Meat (beef)</td>
<td>—</td>
<td>Enrichment and percholl density centrifugation</td>
<td>AmpliTaq® (Applied Biosystems)</td>
<td>—</td>
<td>45</td>
</tr>
<tr>
<td>Meat (pork)</td>
<td>Collagen</td>
<td>DNA extraction (phenol-chloroform)</td>
<td>Taq (Roche Diagnostics)</td>
<td>MgCl₂</td>
<td>5</td>
</tr>
<tr>
<td>Milk</td>
<td>Proteases</td>
<td>DNA extraction (boiling, centrifugation), Chelex -100</td>
<td>Tth® (Roche Diagnostics)</td>
<td>—</td>
<td>50-69</td>
</tr>
<tr>
<td>Soft cheese</td>
<td>—</td>
<td>Aqueous two-phase systems</td>
<td>Taq (Roche Diagnostics)</td>
<td>—</td>
<td>73</td>
</tr>
<tr>
<td>Sugar solution</td>
<td>—</td>
<td>Filtration, cell lysis</td>
<td>Taq (Roche Diagnostics)</td>
<td>—</td>
<td>74</td>
</tr>
<tr>
<td><strong>Clinical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood</td>
<td>—</td>
<td>Dilution, centrifugation, and washing with NaOH</td>
<td>Taq (Roche Diagnostics)</td>
<td>BSA</td>
<td>47,39</td>
</tr>
<tr>
<td>Blood</td>
<td>Hemoglobin, lactoferrin, immunoglobulin G</td>
<td>—</td>
<td>tTth® (Applied Biosystems)</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>Feces</td>
<td>—</td>
<td>DNA extraction (Catrimox-14)</td>
<td>Taq (Takara Shuzo)</td>
<td>Gelatin</td>
<td>75</td>
</tr>
<tr>
<td>Feces</td>
<td>Bilirubin, urobilogens, bile salts, and humic materials</td>
<td>Filtration, DNA extraction (phenol-chloroform), Chelex-100</td>
<td>Tth® (Roche Diagnostics)</td>
<td>—</td>
<td>61</td>
</tr>
<tr>
<td>Feces (porcine)</td>
<td>Polysaccharides, bilirubin, and bile salts</td>
<td>Enrichment and DNA extraction (PrepMan kit [Applied Biosystems])</td>
<td>tTth® (Applied Biosystems)</td>
<td>—</td>
<td>22</td>
</tr>
<tr>
<td>Middle ear effusion</td>
<td>—</td>
<td>DNA extraction (phenol-chloroform)</td>
<td>AmpliTaq Gold® (Applied Biosystems)</td>
<td>—</td>
<td>76</td>
</tr>
<tr>
<td>Skeletal muscle tissue</td>
<td>Myoglobin</td>
<td>RNA extraction (modified acid guanidinium isothiocyanate phenol-chloroform method)</td>
<td>tTth® (Applied Biosystems)</td>
<td>—</td>
<td>77</td>
</tr>
<tr>
<td>Swabs</td>
<td>—</td>
<td>RNA extraction (Qiamp kit [Qiagen])</td>
<td>Tth® (Roche Diagnostics)</td>
<td>—</td>
<td>62</td>
</tr>
<tr>
<td>Tissue</td>
<td>Melanin</td>
<td>RNA extraction (phenol-chloroform)</td>
<td>Red Hot (Advanced Biotechnologies)</td>
<td>BSA</td>
<td>78</td>
</tr>
<tr>
<td>Tissue (paraffin-embedded)</td>
<td>Histological stains, preservatives</td>
<td>DNA extraction (QIA quick gel extraction kit [Qiagen])</td>
<td>AmpliTaq Gold® (Applied Biosystems)</td>
<td>—</td>
<td>63</td>
</tr>
<tr>
<td>Vitreous fluids</td>
<td>—</td>
<td>—</td>
<td>Tth® (Promega Corporation), Tth® (Promega Corporation)</td>
<td>TritonX-100</td>
<td>79</td>
</tr>
</tbody>
</table>

(continued)
### TABLE 5.1 (CONTINUED)
Examples of Pre-PCR Processing Strategies

<table>
<thead>
<tr>
<th>Sample</th>
<th>Recognized Inhibitor&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Sample Preparation</th>
<th>DNA Polymerase</th>
<th>Facilitator&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urine</td>
<td>Urea, beta-HCG, crystals</td>
<td>DNA extraction (PEG)</td>
<td>AmpliTaq (Roche Diagnostics)</td>
<td>—</td>
<td>80, 81</td>
</tr>
<tr>
<td>Environmental</td>
<td></td>
<td>DNA extraction (cetyltrimethylammonium bromide)</td>
<td>Promega Taq (Promega Corporation)</td>
<td>Tween 20</td>
<td>82</td>
</tr>
<tr>
<td>Plant</td>
<td>Acidic polysaccharides</td>
<td>Viruses, centrifugation with PEG, and RNA extraction (RNeasy® plant mini kit [Qiagen])</td>
<td>MuLV, AmpliTaq Gold (Applied Biosystems)</td>
<td>Polyvinylpyrrolidone (PVP), gp32</td>
<td>13, 83</td>
</tr>
<tr>
<td>Sludge</td>
<td>Polyphenols, humic acids, and heavy metals</td>
<td>Enrichment and DNA extraction (QIAamp tissue kit [Qiagen])</td>
<td>Promega Taq (Promega Corporation)</td>
<td>—</td>
<td>84, 85</td>
</tr>
<tr>
<td>Soil</td>
<td>Humic materials</td>
<td>Filtration, enrichment, and DNA extraction (Isoquick DNA extraction [Ocra Research Ltd] or PrepMan kit [Applied Biosystems])</td>
<td>Platinum Taq (Life Technologies)</td>
<td>—</td>
<td>86, 87</td>
</tr>
<tr>
<td>Water</td>
<td></td>
<td>Enrichment</td>
<td>Tth&lt;sup&gt;d&lt;/sup&gt; (Applied Biosystems)</td>
<td>Glycerol</td>
<td>88</td>
</tr>
<tr>
<td>Other</td>
<td>Culture (bacteria)</td>
<td>Enrichment</td>
<td>Tth&lt;sup&gt;d&lt;/sup&gt; (MBI Fermentas)</td>
<td>W1</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Culture</td>
<td>Enrichment</td>
<td>—</td>
<td>detergent, BSA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNA, RNA</td>
<td>Chemical cell lysis</td>
<td>Tth&lt;sup&gt;d&lt;/sup&gt; (Amersham)</td>
<td>—</td>
<td>48</td>
</tr>
</tbody>
</table>

<sup>a</sup> Inhibitor either suggested or identified in the reference.

<sup>b</sup> Pre-PCR processing is converting a complex biological sample containing the target microorganisms into PCR-amplifiable samples and includes sample preparation, DNA polymerases, and the use of facilitators.

<sup>c</sup> Facilitator is a compound that can improve the performance of PCR and which is added extra to the PCR mixture.

<sup>d</sup> The reference uses an alternative DNA polymerase in order to overcome PCR inhibition.
different groups: (1) proteins, (2) organic solvents, (3) nonionic detergents, (4) biologically compatible solutes, and (5) polymers. With the commercial introduction of new DNA polymerases, a number of suppliers have added amplification facilitators to the accompanying buffers (a factor not always considered when the performances of different polymerases were compared).

Amplification facilitators can affect amplification at different stages and under different conditions (see Chapter 6 by Raj Chakrabarti on novel PCR-enhancing compounds and their modes of action). Firstly, amplification facilitators can increase or decrease the thermal stability of the DNA template. For example, organic solvents such as dimethyl sulfoxide (DMSO) and formamide destabilize nucleotide base pairs in solution.65 Some amplification facilitators might affect the error rate of the DNA polymerase. Tween 20 reduces false terminations of the primer extension reaction.66 Finally, facilitators can be used to relieve the amplification inhibition caused by complex biological samples. Many examples (Table 5.1) are known in this group, such as the use of bovine serum albumin (BSA) or the single-stranded DNA-binding protein encoded by gene 32 of bacteriophage T4 (gp32). One of the proposed mechanisms behind the ability of BSA to relieve inhibition is the binding of inhibitory compounds such as blood, heme, and phenolics.67,68 Furthermore, proteins (BSA and gp32) could act as a target for proteases thereby protecting the polymerase or even the DNA itself (gp32).

V. CONCLUDING REMARKS

The pretreatment of a complex biological sample is crucial, and for successful PCR the following requirements have to be fulfilled:71 (1) complete lack or low concentration of PCR-inhibitory components in the sample and (2) sufficient concentration of target DNA. The aim of the pre-PCR treatment is to convert a complex biological sample containing the target microorganisms into PCR-amplifiable samples. Since most complex biological samples contain PCR inhibitors,8 numerous pre-PCR protocols have been developed (Table 5.1). Many PCR protocols combine sample preparation methods from different categories. A common strategy is to combine a pre-enrichment method with a biochemical method13,22 or with a physical sample preparation method.39 In general, RNA/DNA extraction methods provide templates of high quality, but these methods are usually complex. Physical methods are favorable as they do not influence the specificity of the PCR protocol as may the enrichment and immunological methods. The reason for the many PCR protocols and pre-PCR processing methods employed is that the most suitable strategy depends on the nature of the sample and the purpose of the PCR analysis. For instance, various sample preparation methods have been developed to remove or reduce the effects of PCR inhibitors without knowing the identity of the PCR inhibitors and/or understanding the mechanisms of inhibition. Consequently, the conditions for DNA amplification must be optimized by the use of efficient pre-PCR processing. Several different pre-PCR processing strategies can be used: (1) optimization of the sample preparation method, (2) optimization of the DNA amplification conditions by the use of alternative DNA polymerases and/or amplification facilitators, and (3) a combination of both strategies. The growing demand for rapid, robust, and user-friendly PCR protocols means that research in pre-PCR processing is likely to expand in the future. In addition, we are gaining knowledge about the mode of action of different amplification facilitators, as well as the mechanisms by which various factors negatively affect PCR. This will make it easier to choose suitable modifications of the basic PCR master mix that might overcome the problems.

ACKNOWLEDGMENTS

This work was supported by grants from the Swedish Agency for Innovation Systems (VINNOVA), the Foundation of the Swedish Farmers Supply and Crop Marketing Cooperation (SL-stiftelsen), the Swedish Foundation for Strategic Research, and from the Commission of the European Community within the program “Quality of Life and Management of Living Resources,” QLRT-1999.00226.
REFERENCES


57. Sakaguchi, A. Y. et al., Cautionary note on the use of dUMP-containing PCR primers with \( Pfu \) and \( Vent\text{\textregistered} \) DNA polymerases, *BioTechniques*, 21, 368, 1995.

75. Uwatoko, K. et al., Rapid and efficient method to eliminate substances inhibitory to the polymerase chain reaction from animal fecal samples, *Vet. Microbiol.*, 52, 73, 1996.


PCR Optimization
6 Novel PCR-Enhancing Compounds and Their Modes of Action

Raj Chakrabarti

CONTENTS

I. Introduction .......................................................................................................................... 51
II. Materials and Methods ...................................................................................................... 52
   A. PCR Reactions ................................................................................................................ 52
   B. DNA Melting Analysis ................................................................................................... 52
   C. Taq Polymerase Activity and Thermostability Analysis .............................................. 53
III. Results ............................................................................................................................. 53
   A. Templates Used and Definition of Terms ...................................................................... 53
   B. PCR Enhancement by Organic Solvents ........................................................................ 54
   C. DNA Melting .................................................................................................................. 56
   D. Taq Polymerase Thermostability and Activity ............................................................ 57
IV. Discussion ....................................................................................................................... 58
   A. Identification of Outstanding PCR Enhancers .............................................................. 58
   B. Mechanisms of PCR Enhancement .............................................................................. 60
Acknowledgments .................................................................................................................. 62
References .................................................................................................................................. 62

I. INTRODUCTION

DNA amplification by the polymerase chain reaction is often frustrated by inadequate yield of the target DNA sequence and the accompanying amplification of undesired nonspecific products. These problems can be especially severe in the case of targets with high GC content. The complementary strands of such templates have high melting temperatures \( T_m \), which makes them more difficult to denature. Moreover, in their single-stranded form, they are more prone to form secondary structures that impede polymerase extension. A common strategy for the alleviation of these problems has been to include small quantities of certain organic chemicals, such as DMSO, formamide, betaine or glycerol, in the reaction mixture. Recently, extensive structure-activity studies among four groups of compounds — sulfoxides, sulfones, amides and diols — have identified a wide variety of novel PCR enhancers, many of which are significantly more potent than those commonly used.
II. MATERIALS AND METHODS

A. PCR REACTIONS

Polymerase chain reactions were carried out under the following conditions: 10 mM Tris-HCl (pH 8.8), 50 mM KCl, 1.5 mM MgCl₂, 0.01% (w/v) gelatin, 0.2 μM primers, 0.06 ng/μl template, 0.2 mM each dNTP (Stratagene), 0.04 U/μl Taq polymerase (Stratagene). Synthesis of cDNA was carried out using the First-Strand RT-PCR kit (Stratagene) on the respective mRNAs (Clontech). The target sequences and their primers (Genosys) were as follows.

Human myeloid leukocyte c-jun cDNA (996 bp, 64% GC):
primer j1: d(ATGACTGCAAAGATGGAAACG),
primer j2: d(TCAAAATGTTTGCAACTGCG).

Human prostate-specific membrane antigen (PSM) cDNA (511 bp, 52% GC with a 158-bp region containing 73% GC):
primer p1: d(AAACACTGC-TGTGGTGGA),
primer p2: d(TAGCTCAACAGAATCCAGGC).

Bovine brain glycolipid transfer protein (GTP) cDNA (660 bp, 58% GC):
primer g1: d(GAATTCGAAATGGCGCTGCTGG),
primer g2: d(CTCGAGGTCCAGAGTACCGCTGTG).

Bovine brain N-WASP cDNA (1518 bp, 49% GC):
primer N1: d(ATGAGCTCCGGCCAGCAGC),
primer N2: d(TCAGTCTTCCCATTCATCATCATCCTC).

Amplification reactions were carried out on a Robocycler Gradient 96 thermal cycler (Stratagene) using 50-μl solutions in 200-μl thin-walled tubes. Hotstart consisted of heating at 95°C for 5 min, followed by 54°C for 5 min, prior to Taq polymerase addition. Amplifications were run for 30 cycles with 92°C or 95°C denaturation (1 min), 72°C extension (1 min), and annealing at either fixed concentrations of additives, using temperature gradients (44–58°C, 46–60°C, 48–62°C), or at varying concentrations of additives using fixed temperatures (1 min).

Chemicals were from Acros, except for the diols, isobutyramide, and betaine monohydrate (Sigma-Aldrich); the pyrrolidones (BASF); methyl sec-butyl sulfoxide and d-valerolactam (Lancaster Synthesis, Clariant); DMSO and N-formyl pyrrolidine (Fluka); and formamide (Gibco-BRL).

Electrophoresis of amplification products was done on 0.8% agarose gels in which 20 μl reaction products were loaded with 4-μl loading buffer. Gels were run at 60 V for 80 min and stained with ethidium bromide. Amplification products were quantitated using an ImageScanner densitometer and ImageMaster Total Lab software (Amersham). Interpolation of data points at intermediate concentrations was carried out using the method of cubic splines.15

B. DNA MELTING ANALYSIS

DNA melting temperatures and thermodynamic parameters were determined by using a fluorescence resonance energy transfer (FRET)-based method.14,16,17 Two pairs of complementary 40-mer DNAs with the following sequences were synthesized: oligo A1: d(GACTGCAAAAGATGGAACGACCTTTCTATGACGATGCGGCTC); oligo A2: d(GAGGGCATCGTCATAGAAGGTCGTTTCCATCTTTGCAGTC); oligo B1: d(CTGGCTGTGCGATGGGGCGCTGGTGACGCGGGGTGGCTTC); oligo B2: d(GAAGCCACCCCGCTGCACGCCCAGGGCGCCACAGCAGCAG). A1 and B1 were 5’-modified with fluorescein, and A2 and B2 were 3’-modified with carboxytetramethylrhodamine.
Novel PCR-Enhancing Compounds and Their Modes of Action

(TAMRA). All melting studies were carried out on solutions containing the complementary oligonucleotides at concentrations of either 5 or 100 nM in 50 mM NaCl and 10 mM phosphate buffer, pH 7.3. Temperature was ramped from 25 to 92°C at a rate of 10°C per h and fluorescein fluorescence values (494 nm excitation, 520 nm detection) were recorded at 0.1°C increments using a SPEX Fluoromax-3 fluorometer with Peltier temperature controller.

C. Taq DNA Polymerase Activity and Thermostability Analysis

Taq DNA polymerase activity was assayed with a nonradioactive fluorescence-based method using the PicoGreen dsDNA quantitation reagent.\textsuperscript{14,18,19} Briefly, 2.6 µg M13 17-mer primer (New England Biolabs) mixed with 100 µg M13 mp18 (+)-strand DNA (Amersham) in 1X Taq polymerase buffer was heated to 70°C for 5 min and cooled slowly at room temperature for 15 min (primer-template solution); dNTPs were then added at a concentration of 670 µM each, keeping the buffer concentration constant and making the template concentration 0.067 µg/µl. 15 µl of this solution was added to 33 µl of appropriate concentrations of cosolvent in 200 µl thin-walled tubes, the tubes preincubated at 72°C for 1 min and 2 µl of Taq polymerase solution (1 U/µl concentration) added. Reactions were stopped after 15-min incubation with 5 µl of 0.25 M EDTA (Sigma), and dsDNA concentrations quantitated using the PicoGreen reagent (Molecular Probes) according to the instructions of the manufacturer (480 nm excitation/524.5 nm detection, in a SPEX Fluoromax-3 fluorometer).

The thermostability of Taq polymerase was also assayed using a PicoGreen-based method. First, 23 µl of the appropriate concentrations of cosolvents in 1X Taq reaction buffer were incubated at 92°C or 95°C for 1 min in 200 µl thin-walled tubes, after which 2 µl of 1.5 U/µl Taq polymerase was added. At incremental times, tubes were removed and immediately chilled on ice. The primer-template solution was diluted to a concentration of 0.042 µg/µl template in Taq reaction buffer and 24 µl of this solution added to the chilled tubes. Within 15 min, residual activity was determined at 72°C using the assay described above, except reactions were now initiated by the addition of 1 µl of a solution containing 10 mM of each dNTP. The resulting fluorescence values were compared to those of controls containing the same cosolvent concentrations that were carried through identical steps with the exception of thermal incubation.

III. RESULTS

A. Templates Used and Definition of Terms

The effects of cosolvents on PCR amplification were studied using four templates with differing GC content — c-jun, PSM, GTP, and N-WASP (see Section 11, Materials and Methods). Amplification of c-jun was impossible in the absence of PCR enhancers, even using denaturation temperatures as high as 96°C. Amplification of PSM is weak and accompanied by the amplification of a particularly prominent low molecular weight nonspecific band, most likely the result of strong secondary structure in its 73% GC segment. GTP and N-WASP amplification are also weak and nonspecific.\textsuperscript{11}

For each family of compounds, one of these four templates was chosen as the primary subject of the investigation with which each additive was tested. A subgroup of functional additives was then tested with additional templates. The following experimental strategy was used to characterize the PCR-enhancing capabilities of the various cosolvents. First, each of the additives in a given family was screened with the primary template at a few disparate concentrations over an annealing temperature gradient, using a denaturation temperature of either 92°C or 95°C to determine the optimal annealing temperature. If the same number of compounds in the family was effective at both 92°C and 95°C, 92°C was used as the denaturation temperature for all subsequent experiments since the differentiation in additive performance was more pronounced at this temperature; otherwise, 95°C was used. The functional additives were then tested over molar concentration gradients at their optimal annealing temperatures.
Description of the effectiveness of the various additives was achieved in the case of each target through the assignment of three quantities, termed potency, specificity, and effective range to each compound. The potency of an additive is defined as the maximum densitometric volume of target band amplification over the concentration gradient tested for that additive. Maximal target band volumes were interpolated from the data sets by fitting cubic splines to the data. Potencies of the various compounds tested in each family were normalized to those of the standard compounds for that family (formamide for amides, DMSO for sulfoxides and sulfones, and glycerol for diols), which were assigned values of 1. The specificity of an additive at a particular concentration is defined as the ratio of the volume of target band amplification to the total volume of all bands, including undesired nonspecific bands, expressed as a percent. The definition of effective range evolved during the course of our investigations. For initial studies with low molecular-weight amides, the effective range was defined in terms of an additive’s leveling-off concentration (the approximate concentration at which product yield was not increased any further) and its cutoff concentration (above which amplification was inhibited). In subsequent studies, the effective range was more quantitatively defined as the interval of concentrations, determined from the additive’s cubic spline curve over which the volume of target band amplification is at least 50% of the maximal volume.

B. PCR Enhancement by Organic Solvents

Table 6.1 and Table 6.2 show the compounds that we examined in our studies and summarize their effects. The amide and sulfoxide cosolvent families were chosen because formamide and DMSO, the simplest members of these groups, are the most widely used PCR enhancers. Sulfones, similar to sulfoxides with the exception that the sulfur atom is double-bonded to two oxygen atoms instead of just one, were studied based on our findings that a number of sulfoxides are effective in PCR enhancement. Diols were tested because monools, though good DNA denaturants, also denature enzymes and glycerol (a triol), while a stabilizer of protein structure is only a weak DNA denaturant and hence affords only marginal PCR enhancement.

<table>
<thead>
<tr>
<th>Additive</th>
<th>Potency</th>
<th>Specificity</th>
<th>Effective Range (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Leveling-Off</td>
</tr>
<tr>
<td>2-Pyrrolidone</td>
<td>1.18</td>
<td>86%</td>
<td>0.12</td>
</tr>
<tr>
<td>HEP</td>
<td>0.98</td>
<td>79%</td>
<td>0.08</td>
</tr>
<tr>
<td>NMP</td>
<td>1.10</td>
<td>81%</td>
<td>0.15</td>
</tr>
<tr>
<td>Formamide</td>
<td>1.00</td>
<td>87%</td>
<td>0.67</td>
</tr>
<tr>
<td>MMF</td>
<td>1.07</td>
<td>79%</td>
<td>0.51</td>
</tr>
<tr>
<td>DMF</td>
<td>0.98</td>
<td>77%</td>
<td>0.41</td>
</tr>
<tr>
<td>Acetamide</td>
<td>1.07</td>
<td>98%</td>
<td>0.25</td>
</tr>
<tr>
<td>MMA</td>
<td>0.81</td>
<td>85%</td>
<td>0.27</td>
</tr>
<tr>
<td>DMA</td>
<td>0.88</td>
<td>80%</td>
<td>0.29</td>
</tr>
<tr>
<td>Propionamide</td>
<td>1.13</td>
<td>82%</td>
<td>0.21</td>
</tr>
<tr>
<td>Isobutyramide</td>
<td>1.10</td>
<td>85%</td>
<td>0.23</td>
</tr>
<tr>
<td>Control</td>
<td>negligible</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

*a Normalized densitometric volume of target band averaged over effective range (formamide = 1).

*b Represents best specificity (densitometric volume of target as percent of total volume) over the effective range.
TABLE 6.2
Potency, Specificity, and Effective Range of PCR-Enhancing Additives

<table>
<thead>
<tr>
<th>Additive</th>
<th>c-Jun (996 bp; 64%) Potency(^{b})</th>
<th>GTP (660 bp; 58%) Potency</th>
<th>PSM (511 bp; 52%) Potency</th>
<th>Effective Range (M)(^{d})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMSO</td>
<td>1.00</td>
<td>1.00*</td>
<td>89%*</td>
<td>1.00*</td>
</tr>
<tr>
<td>Propyl sulfoxide</td>
<td>0.25</td>
<td>2.82*</td>
<td>100%*</td>
<td>0.21*</td>
</tr>
<tr>
<td>Methyl butyl sulfoxide</td>
<td>3.33*</td>
<td>100%*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetramethylene sulfoxide</td>
<td>1.18</td>
<td>3.68*</td>
<td>100%*</td>
<td>1.74*</td>
</tr>
<tr>
<td>Methyl sulfone</td>
<td>0.85</td>
<td>1.06*</td>
<td>100%*</td>
<td>1.40*</td>
</tr>
<tr>
<td>Ethyl sulfone</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propyl sulfone</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sulfolane</td>
<td>1.10</td>
<td>3.30*</td>
<td>100%*</td>
<td>1.64*</td>
</tr>
<tr>
<td>3-Sulfolene</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formamide</td>
<td>1.00</td>
<td>1.00</td>
<td>100%</td>
<td>1.00</td>
</tr>
<tr>
<td>Acetamide</td>
<td>0.99</td>
<td>0.53</td>
<td>100%</td>
<td>0.95</td>
</tr>
<tr>
<td>2-Pyrrolidone</td>
<td>2.90</td>
<td>1.37</td>
<td>100%</td>
<td>2.62</td>
</tr>
<tr>
<td>δ-Valerolactam</td>
<td>3.04</td>
<td>1.48</td>
<td>100%</td>
<td>4.37</td>
</tr>
<tr>
<td>ε-Caprolactam</td>
<td>1.09</td>
<td>1.52</td>
<td>100%</td>
<td>2.84</td>
</tr>
<tr>
<td>2-Azacyclooctanone</td>
<td>2.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formyl pyrroline</td>
<td>1.81</td>
<td>1.42</td>
<td>100%</td>
<td>1.09</td>
</tr>
<tr>
<td>Formyl piperidine</td>
<td></td>
<td>1.96</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Formyl morpholine</td>
<td>4.53</td>
<td>1.72</td>
<td>100%</td>
<td>1.48</td>
</tr>
<tr>
<td>Glycerol</td>
<td>1.00</td>
<td>1.00</td>
<td>100%</td>
<td>1.00</td>
</tr>
<tr>
<td>Ethylene glycol</td>
<td>4.46</td>
<td>2.21</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>1,2-Propanediol</td>
<td>4.61</td>
<td>2.28</td>
<td>100%</td>
<td>3.41</td>
</tr>
<tr>
<td>1,3-Propanediol</td>
<td>1.92</td>
<td>1.39</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>1,2-Butanediol</td>
<td>4.92</td>
<td>2.03</td>
<td>100%</td>
<td>1.91</td>
</tr>
<tr>
<td>1,3-Butanediol</td>
<td>5.92</td>
<td>3.03</td>
<td>100%</td>
<td>4.55</td>
</tr>
<tr>
<td>1,4-Butanediol</td>
<td>6.31</td>
<td>3.11</td>
<td>100%</td>
<td>2.96</td>
</tr>
<tr>
<td>2,4-Pentanediol</td>
<td></td>
<td>2.18</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>1,5-Pentanediol</td>
<td>7.69</td>
<td>3.57</td>
<td>100%</td>
<td>2.46</td>
</tr>
<tr>
<td>cis-1,2-Cyclopentanediol</td>
<td>3.25</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,6-Hexanediol</td>
<td>2.86</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPD</td>
<td>1.17</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Betaine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: All data obtained under 95°C denaturing conditions, unless marked by an asterisk (92°C). Please note that potencies can be directly compared only when the compounds have been normalized to the same standard under the same denaturation conditions.

\(^{a}\) The PSM fragment includes a 158-bp region with 73% GC.

\(^{b}\) Normalized densitometric volume of target band: standard compound (sulfoxides, sulfones, betaine: DMSO; amides: formamide; diols: glycerol). Control potencies (no additive): GTP (95°C) ≤ 25% of standards, (92°C) ≤ 5% standards; PSM (95°C) ≤ 65% standards, PSM (92°C) ≤ 20% standards; c-jun (92°C/95°C) = 0.

\(^{c}\) Best specificity (densitometric volume of target as percent of total volume) over the effective range. Control specificities (no additive): GTP (95°C) = 20%, (92°C) = 53%; PSM (95°C) = 29%, (92°C) = 34%; c-jun (92°C/95°C) = 100%.

\(^{d}\) Concentration interval over which densitometric volume of target is at least 50% of maximum volume; data shown represent the broadest range obtained over all three templates.
The amide additives tested can be divided into four subclasses: the formamides, the acetamides, the higher chain primary amides, and the cyclic pyrrolidones.  

N-W ASP was used as the primary target in the study of low molecular-weight amides. With the exception of N-cyclohexyl pyrrolidone, all of the amides were found to be effective PCR enhancers. Potency, specificity, and effective range data for the various amides with N-W ASP are provided in Table 6.1. The compounds that performed best with N-W ASP, 2-pyrrolidone and acetamide, and the standard additive formamide were tested further with the higher GC-content templates c-jun, PSM, and GTP. The potencies and specificities exhibited with these templates are listed in Table 6.2.

Given the outstanding performance of 2-pyrrolidone, we decided to investigate the PCR-enhancing properties of larger cyclic amides. In addition to endocyclic amides such as 2-pyrrolidone, in which the amide carbonyl group is contained within the ring structure, we included in this study exocyclic amides where the carbonyl group is located outside the ring. GTP was used as the primary template in the cyclic amide investigation; c-jun and PSM were studied less extensively. Performance data can be found in Table 6.2, except for 2-azacyclononanone and N-acetyl morpholine, which were ineffective.

All known water-soluble symmetrical n-alkyl sulfoxides were tested, with the exception of ethyl sulfoxide, which was not readily available from known chemical sources. Methyl sec-butyl sulfoxide, which has a total number of carbons between those of ethyl and propyl sulfoxides, was used instead. Tetramethylene sulfoxide, a simple cyclic sulfoxide, was also included in the study. Potency, specificity, and effective range with GTP as the primary template are listed in Table 6.2; butyl sulfoxide was ineffective.

The group of sulfones was selected to include the simplest n-alkyl sulfones, a simple cyclic sulfone (sulfolane), a ring-substituted cyclic sulfone (2,4-dimethylsulfolane), and a cyclic alkene sulfone (sulfolene). Because of their low water solubilities, N-butyl and higher alkyl sulfones were omitted and 2,4-dimethylsulfoxide was ineffective. We used c-jun as the primary template for the sulfone investigation (Table 6.2).

The chosen diols span the spectrum of low-molecular weight compounds from this family. For every chain length, multiple structural isomers were tested in order to determine the effects of relative hydroxyl location on PCR amplification. The cyclic diols cis-1,2-cyclopentanediol and trans-1,2-cyclopentanediol were examined to study simultaneously the effects of hydroxyl conformational restriction and cyclic structure on amplification. 2-methyl-2,4-pentanediol (MPD), a hexanediol structural isomer, was included to examine the effects of chain branching and to determine whether this known stabilizer of certain protein and DNA conformations exhibits any unique properties. Glycerol was included as a reference standard and GTP was chosen as the primary template for the diol studies (see Table 6.2 for a summary of the results). Found to be ineffective were 1,2-pentanediol, 1,2-hexanediol and trans-1,2-cyclopentanediol.

Figure 6.1A displays the variation of c-jun amplification with concentration of selected diols and glycerol, along with the accompanying cubic spline curve fits. The gel electrophoresis results from a representative diol concentration gradient, that of 1,2-propanediol with c-jun, can be found in Figure 6.1B. Figure 6.1C compares the best results obtained in the enhancement of c-jun amplification with selected diols and glycerol.

For the sake of comparison, betaine, which has recently been described as a particularly potent enhancer of high-GC template amplification was included in the investigation (Table 6.2).

C. DNA Melting

DNA melting curves were recorded in the presence of varying concentrations of selected additives. The majority of these solvents were tested with both 50% GC and 73% GC 40-mers, to determine effect of base composition on the modulation of DNA melting. In all cases, the melting temperatures (Tm) of the DNAs were depressed by the presence of the solvents (Table 6.3). The decrease in Tm with increasing additive concentration was found to be approximately linear for each cosolvent.
within the concentration range that depressed the \( T_m \) by < 10°C. \( \Delta H \) was obtained from the slope of a plot of \( 1/T_m \) vs. \( \ln C_t \) (total strand concentration), where \( C_t = 10 \text{ nM} \) and 400 nM, and \( \Delta S \) was obtained from the y-intercept. \( \Delta H \) and \( \Delta S \) in the absence of solvents were first calculated from the control melting curve of the 73% GC oligomer. Molar \( \Delta \Delta H \) and \( \Delta \Delta S \) elicited by the cosolvents were then determined by comparing the melting curves for this oligomer in the presence of the solvents to the control. As can be seen from the equation \( \Delta \Delta S = \Delta H \times (1/T_m) + (1/T_m) \times (\Delta H) \), which is the differential form of \( \Delta G = \Delta H - T \Delta S \) at the melting temperature, \( \Delta \Delta S \) can be decomposed into cooperative (\( \Delta \Delta H \)-dependent) and noncooperative (\( \Delta T_m \)-associated) parts. \( \Delta \Delta H \), fractional \( \Delta \Delta S \), and \( \Delta \Delta S_{\text{noncoop}} \) per molar concentration of the solvents are also presented in Table 6.3.

D. Taq Polymerase Thermostability and Activity

We studied the kinetics of thermal deactivation of Taq DNA polymerase by heating the enzyme in buffer with selected additives either to 92°C or 95°C for various times, cooling and measuring
residual activity at 72°C. In the majority of cases, additive concentrations near the optimal PCR concentrations were tested. The deactivation data was fit with simple exponential decay curves. The corresponding first-order decay rate constants are listed in Table 6.4. The activity of Taq polymerase was measured at 72°C in the presence of various concentrations of selected cosolvents. In each case, the decline in activity with solvent concentration was roughly linear until a low level of activity was reached, at which point the rate of decline decreased in most cases. The approximate rates of decline in activity with concentration are presented in Table 6.4 along with the range of concentrations over which the R factor for linear regression was at least 0.97. The y-intercepts of the linear fits are also included. In the cases of MPD and 1,6-hexanediol, the y-intercepts are particularly low; the activity dropped sharply between 0 and 0.1 M concentration, at which point it began to decay in an approximately linear fashion. In all cases, the reduction in activity elicited by cosolvents was fully reversed upon dilution with buffer.

### IV. DISCUSSION

#### A. IDENTIFICATION OF OUTSTANDING PCR ENHANCERS

The results in Table 6.1 and Table 6.2 clearly establish that various amides, sulfoxides, sulfones, and diols are capable of improving PCR yield and specificity, often dramatically. A wide variety of compounds from each family are effective. In many cases, the potencies and specificities of these cosolvents are considerably greater than those of standard PCR enhancers, including betaine.

While virtually every amide tested was effective with the moderately high-GC template N-WASP, the performance of the cyclic amide 2-pyrrolidone was outstanding with the three higher GC targets. Subsequent studies with larger endo- and exocyclic amides confirmed that the cyclic amides are superior PCR enhancers. We identify N-formyl morpholine as the most effective cyclic amide in the enhancement of PCR amplification. Overall, its range, potency, and specificity were the greatest of all the amides examined. Another highly effective enhancer is δ-valerolactam; along with N-formyl morpholine, it was the only larger cyclic amide that performed well across all three

### TABLE 6.3

<table>
<thead>
<tr>
<th>Additive</th>
<th>(-dT_m/dC_A) (K/M)</th>
<th>(50%) GC</th>
<th>(73%) GC</th>
<th>(\Delta H/dC_A) (kJ/M)</th>
<th>(73%) GC</th>
<th>(\Delta S/dC_A) (J/mol KM)</th>
<th>(73%) GC</th>
<th>(\Delta S/dC_A),noncoop (J/mol KM)</th>
<th>(73%) GC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2-Propanediol</td>
<td>1.6</td>
<td>1.6</td>
<td>42</td>
<td>138</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,3-Propanediol</td>
<td>0.9</td>
<td>0.9</td>
<td>0</td>
<td>84</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,2-Butanediol</td>
<td>8.4</td>
<td>7.4</td>
<td>0</td>
<td>42</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,3-Butanediol</td>
<td>3.9</td>
<td>3.7</td>
<td>0</td>
<td>64</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,4-Butanediol</td>
<td>6.1</td>
<td>5.6</td>
<td>0</td>
<td>167</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,4-Pentanediol</td>
<td>NA</td>
<td>5.0</td>
<td>51</td>
<td>201</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,5-Pentanediol</td>
<td>19.4</td>
<td>14.7</td>
<td>0</td>
<td>167</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cis-1,2-Cyclopentanedi</td>
<td>26.1</td>
<td>18.6</td>
<td>0</td>
<td>218</td>
<td>218</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPD</td>
<td>3.7</td>
<td>3.5</td>
<td>265</td>
<td>751</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,6-Hexanediol</td>
<td>27.3</td>
<td>19.1</td>
<td>0</td>
<td>218</td>
<td>218</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sulfolane</td>
<td>NA</td>
<td>13.5</td>
<td>-102</td>
<td>-134</td>
<td>154</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetramethylene sulfide</td>
<td>NA</td>
<td>20.2</td>
<td>-138</td>
<td>-161</td>
<td>230</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: K = degrees kelvin; C_A = concentration of additive; M = molar; kJ = kilojoules; % GC designates GC content of the DNA used; accuracy of reported thermodynamic values is ±10 kJ/mol; \(\Delta H_{\text{control}} = 1410\) kJ/mol.
Novel PCR-Enhancing Compounds and Their Modes of Action

templates. Somewhat less effective than δ-valerolactam is 2-pyrrolidone; 2-Azacyclooctanone and N-formyl piperidine are both highly potent enhancers but may have ranges that are too narrow for convenient application. Like all cyclic amides, ε-caprolactam and N-formyl pyrrolidine are powerful enhancers of PCR, but their performance is less predictable than those of the most effective compounds. Cyclic compounds within the sulfoxide and sulfone families were found to exhibit superior PCR enhancement as well, especially tetramethylene sulfoxide and sulfolane.

Among the acyclic amide enhancers, acetamide is the most effective. It gives good amplification of the target band and the highest specificity and has an effective range near the top of the list. Methyl sulfone is the best-performing acyclic sulfone. Based on the collective results, it seems to offer an advantage over DMSO with respect to GC-rich targets.

Essentially all low molecular-weight diols are significantly more effective in PCR enhancement than glycerol, the state-of-the-art alcohol enhancer. With templates of low to moderate secondary structure (c-jun and GTP), the effectiveness of diol enhancers appears to improve with increasing chain length up to C₅; however, the relative location of the hydroxyl groups plays an

<table>
<thead>
<tr>
<th>Additive</th>
<th>Thermostability</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Optimal conc. (M)</td>
<td>k₀SC</td>
</tr>
<tr>
<td>None</td>
<td>NA</td>
<td>0.046</td>
</tr>
<tr>
<td>Ethylene glycol</td>
<td>1.2-Propandiol</td>
<td>1.4</td>
</tr>
<tr>
<td>1,2-Butanediol</td>
<td>1.5-Butanediol</td>
<td>0.5</td>
</tr>
<tr>
<td>1,4-Butanediol</td>
<td>1.6-Butanediol</td>
<td>0.8</td>
</tr>
<tr>
<td>1,4-Pentanediol</td>
<td>cis-1,2-Cyclopentanediol</td>
<td>0.6</td>
</tr>
<tr>
<td>MPD</td>
<td>1,5-Pentanediol</td>
<td>0.3</td>
</tr>
<tr>
<td>1,6-Hexanediol</td>
<td>2-Pyrrolidone</td>
<td>0.2</td>
</tr>
<tr>
<td>δ-Valerolactam</td>
<td>2-Valerolactam</td>
<td>0.2</td>
</tr>
<tr>
<td>ε-Caprolactam</td>
<td>2-Azacyclooctanone</td>
<td>0.1</td>
</tr>
<tr>
<td>Formyl pyrrolidine</td>
<td>2-Azacyclooctanone</td>
<td>0.1</td>
</tr>
<tr>
<td>Formyl piperidine</td>
<td>Formyl pyrrolidine</td>
<td>0.1</td>
</tr>
<tr>
<td>Formyl morpholine</td>
<td>Formyl morpholine</td>
<td>0.3</td>
</tr>
<tr>
<td>Sulfolane</td>
<td>2-Azacyclooctanone</td>
<td>0.4</td>
</tr>
<tr>
<td>Tetramethylene sulfoxide</td>
<td>Tetramethylene sulfoxide</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Note: Thermodeactivation rate constants (k) at concentrations of additives near their optimal PCR concentrations; rates of polymerase activity decay at 72°C as functions of additive concentration.

¹ Optimal PCR concentrations can change with the denaturation temperature; the concentrations listed are near the optimal values determined at 95°C for diols and amides, and 92°C for sulfoxides and sulfones, for the various templates studied.

² Fractional change in activity of polymerase (relative to control) per molar concentration of additive (slope of linear least squares fit).

³ Y-intercept of linear least squares fit to activity data.

⁴ Concentration range over which regression coefficient of linear least squares fit is > 0.97.
important role in determining performance. The best-performing additive in the case of these targets is 1,5-pentanediol.

B. MECHANISMS OF PCR ENHANCEMENT

In order to delineate the modes of action of these novel cosolvents, we examined the effects of selected additives on the most relevant mechanistic parameters — DNA melting, polymerase thermostability, and polymerase activity. All of the cosolvents examined depress the melting temperature of DNA (Table 6.3). The absence of any meaningful ΔH for the majority of diols implies that their effect is primarily entropic. The most obvious candidate for a noncooperative positive ΔS is increased solvation of exposed hydrophobic bases in the single-stranded coil state. Tetramethylene sulfoxide is more hydrophobic than sulfolane, implying greater base solvation. Sulfolane and tetramethylene sulfoxide both display a meaningful negative dΔH/dC_A. This effect is most likely due to displacement of weakly bound groove hydrate by the solvent molecules, accompanied by hydrogen bonding of the solvent sulfonyl oxygens with H-bonding moieties in the grooves. The greater depression of ΔH by tetramethylene sulfoxide may be attributable to its superior steric complementarity to these H-bonding moieties.

The roughly monotonic decrease in enzyme activity with concentration observed here (Table 6.4) has some precedent. In these cases, the decline of activity with concentration was found to correlate very closely with a gradual reversible conformational unfolding of the enzymes as monitored by tryptophan fluorescence or CD. The linear rate of the decay of polymerase activity with the concentration of additive implies that monotonic unfolding alone probably dominates in producing the decrease in activity. All of the cosolvents examined decrease the thermostability of the polymerase (Table 6.4). In a more detailed analysis, we have shown that all of these compounds increase the ΔH^* of the thermodeactivation transition state but increase the ΔS^* to an extent that more than compensates for this effect. The different trends in the thermal decay rate constant (a function of ΔG^* = ΔH^* - TΔS^*) at 92°C and 95°C originate in the temperature dependence of the entropic contribution to the free energy.

PCR amplification has been modeled theoretically under a number of simplifying assumptions. This expanded model indicates that both activity decline and thermodegradation contribute to decrease in product yield at high additive concentrations by increasing the minimum extension time (τ) necessary to completely replicate all template strands. In the case of activity decline, the induced increase in τ is comparatively constant over the course of many cycles, whereas thermodegradation causes a continuous increase in τ throughout the reaction. Therefore, differences in activity are unlikely to impact yield (potency) at the intermediate additive concentrations that produce the greatest amplification. Because thermal decay is exponential, and since 30 PCR cycles are used, the model predicts that even relatively benign effects on thermostability will eventually manifest an observable impact on yield. As a result, differences in polymerase thermostability in the presence of different additives are expected to play a role in determining the relative order of their potencies.

On the other hand, the model predicts that activity decline will be the primary cause of the cutoff limiting the effective range. Since this effect is exponentially compounded, once a concentration is reached where this effect becomes important, further increases in concentration will rapidly diminish yield, eventually extinguishing it when individual strands cannot be replicated. By contrast, the significant excess of polymerase used in PCR reactions implies that the effect of thermal decay will only become significant at later cycles after some degree of amplification has already been achieved.

DNA melting temperature depression increases the concentration of single strands at the start of each cycle, increasing the expected yield. This effect is compounded exponentially over all cycles, unlike thermodegradation. Hence, the model predicts that while both melting temperature depression...
and thermodegradation contribute to additive potency. \( T_m \) depression is the primary causative factor determining the relative potencies of additives. Figure 6.2 schematically depicts where \( T_m \) depression, activity decline, and thermodegradation have their most important effects on amplification.

Comparison of the diol potency and effective range data with their mechanistic effects (Table 6.2 to Table 6.4) reveals that these predictions are indeed borne out in the majority of cases. The most striking correlation between observation and prediction is found in the case of range. As can be seen, in virtually every case the cutoff in range occurs at a concentration close to that at which polymerase activity falls to negligible levels. The order of effective ranges of the compounds is the same as the order of their slopes of activity decay, even though these slopes are quite close in some cases. The relative order of potencies of the diols with \( c-jun \) and GTP also correlate well with the order of their effects on thermostability and DNA melting, and the differences observed between \( c-jun \) and GTP can be attributed to the GC dependence of \( T_m \) depression.\(^{14}\) We attribute the aberrant trend observed with PSM to the effects of the additives on this template’s strong secondary structure.
The specificities achieved with the various additives are also directly related to their impacts on secondary structures, which are currently under investigation.

In conclusion, we have identified and characterized a wide variety of novel PCR-enhancing cosolvents which together constitute a library of chemicals for PCR optimization. In many cases, it has been possible to qualitatively correlate the yield and effective concentration range of these compounds with their impact on the reaction. Further mechanistic studies, in conjunction with theoretical models of PCR amplification modified for the presence of cosolvents, should lead to quantitative chemical PCR optimization.\(^\text{14}\)

**ACKNOWLEDGMENTS**

The work summarized herein was supported by the National Science Foundation in the form of a 3-year research fellowship and by the National Institutes of Health (grant GM 44038).

**REFERENCES**


7 Optimizing PCR with the Aid of Experimental Design

Thomas Weissensteiner

CONTENTS

I. Introduction................................................................................................................. 65
II. Primer Design............................................................................................................. 66
III. Optimization of Reaction Mixture and Temperature Cycle........................................ 68
   A. Materials and Methods .......................................................................................... 68
      1. Reagents and Instrumentation ........................................................................ 68
      2. PCR Optimization Spreadsheets .................................................................... 69
      3. Quantifying PCR Yield ................................................................................... 71
   B. Examples and Results ............................................................................................ 71
      1. Step 1: Taguchi Method .................................................................................. 71
      2. Step 2: 5 x 5 Array ......................................................................................... 71
   C. Discussion............................................................................................................... 73
References ................................................................................................................... 75

I. INTRODUCTION

PCR optimization can be considered a two-step process: first comes the design of optimal primer sequences. Their hybridization behavior and priming efficiency, as well as the template sequences inside and outside the amplified target will determine, to a large part, the optimum value and range of the other parameters, e.g., components of the reaction mixture and temperature cycle steps.

Once the primer sequences have been chosen, annealing temperature and/or magnesium ion concentration are generally considered the most critical parameters for PCR success. In addition, variation of primer concentration and dsDNA destabilizing additives may be necessary for some PCR systems to produce satisfactory results. Importantly, all of these variables affect dsDNA stability, a factor that is critical for PCR success in several ways: primer annealing is required to initiate the cycle, but denaturation of template dimers, hairpins, etc. facilitates primer extension. Moreover, generation of optimal yields of a specific product depends on the relative efficiencies of annealing and extension of wanted vs. unwanted amplimers. One might, therefore, predict different but inter-dependent effects for annealing temperature, dsDNA destabilizers, magnesium-ion, and primer concentrations on PCR success. Consequently, the popular strategy of optimizing each factor sequentially may yield local, rather than absolute optima. A safe way to achieve the latter would be testing all possible combinations of all values for all factors (i.e., 81 reactions for 4 factors with 3 values each).

Fortunately, there is a certain amount of redundancy in such a setup and experimental designs have been developed that reduce the number of actual tests while still providing an overall picture of the most influential parameters and their optimum ranges. But although two adaptations of an experimental design (Taguchi) for PCR optimization were published,¹² and one for ELISAs,³ several drawbacks may have hampered their wider use. Planning and analysis required complicated
calculations but no software was provided; the setup was cumbersome and error-prone, involving a high number of pipetting steps. Most importantly, reduced redundancy also meant reduced sensitivity and robustness to experimental error, which if not recognized, can seriously confound the results. In the second part of this chapter, I will present a streamlined protocol with increased sensitivity and reliability, which still keeps the number of test reactions within a manageable range. Spreadsheets for planning and analyzing the experiments are included and can be downloaded from http://www.crcpress.com/e_products/downloads/download.asp?cat_no=1184. Interestingly, the yields of PCRs that were already optimized for MgCl₂ and annealing temperature increased further when adding betaine together with higher primer concentrations.

II. PRIMER DESIGN

Careful choice of the primers can save time-consuming optimization later on. The points to consider include:

- Uniqueness of the primer sequence within the DNA of the sample
- Primer length, melting temperature (Tₘ), and general sequence composition
- 3’-end sequence and 3’ (mis)match
- Secondary structure in the amplified region
- Postamplification procedures

Primers of between 18 and 24 bases stand a good chance of being sequence specific, provided that the annealing temperature is optimal. Ants Kurg et al. even recommend 15 to 18 bp for the amplification of genomic DNA when combined with careful primer design involving searches for false priming sites in the relevant sequence databases. Their chapter contains a detailed discussion of this aspect of primer design which I will not discuss here.

Both primers should be designed to have similar melting temperatures: primer pairs with mismatched Tₘ are problematic because the primer with the higher Tₘ will misprime at lower temperatures, and the primer with the lower Tₘ may not work at higher temperatures. In general, primers with a melting temperature of 55°C to 72°C result in the highest specificity and yield. Among primers with the same Tₘ, long primers with a low GC content tend to produce lower yield, compared to shorter primers with a higher percent GC. It is, therefore, often recommended that both primers should be 50% GC, or at least have similar GC content. However, this does not matter too much, as long as the 3’ end is neither too “loose” or too “sticky,” e.g., the last 7 bases are about 50% GC, with 1 to 2 guanosines or cytidines among the last three.

Single mismatches that are located 4 or more bases from the 3’ end of the primer tend to have little effect on PCR results and can thus be used to introduce restriction sites for cloning (see below). However, primers that are mismatched in the penultimate position are extended poorly by Taq polymerase, and hardly at all when mismatched at the 3’ base. The latter effect can be put to good use, for example, for typing single nucleotide variation (using a nonproofreading polymerase). However, one should note that the ability of Taq polymerase to discriminate such differences depends a lot on the type of mismatch (Table 7.1). A mismatched position on the template followed immediately by a matched base can also reduce specificity through “slipping” of the primer 3’ end, especially when the latter is low in GC content.

It is critical that the forward and reverse primer do not form stable “primer dimers” or hairpins with extendable 3’ ends. This can be checked by visual inspection or with one of many available computer programs, including free software such as PrimerDesign (http://www.chemie.uni-marburg.de/~becker/pdhome.html). For example, a primer with a Tₘ of 60°C or less should not form 3’ dimers with a free energy (ΔG) of less than –3.5 kcal/mol, meaning that two matched 3’ Gs or Cs are permissible if preceded by at least 3 unmatched bases.
If possible, template regions with a high denaturation temperature should be avoided. Some secondary structures of ssDNA, such as tetraplexes formed by long stretches of guanosines \(^4\) stem loops formed by palindromic sequences, \(^5,6\) and triplet repeats \(^7,8\) also cause difficulty. Both types of inhibitory structures depend on template regions with a high GC content, and the average GC content is often used as a first check for how refractory a DNA fragment will be to amplification. If the GC content should vary significantly within the target sequence, however, a plot of the \(T_m\) of individual base pairs vs. their position will be more informative: regardless of the average GC content and \(T_m\), a region of 25 bp where the local \(T_m\) stays at 80°C or above is bound to spell trouble while longer stretches of sequence melting at 70°C seem to be okay. \(^9,10\)

Some applications, such as where only a peptide sequence of the gene of interest is known, require the design of “degenerate primers” covering all possible nucleotide translations. A useful and freely available software designed for this purpose is CODEHOP (http://blocks.fhcrc.org/blocks/codehop.html).

Postamplification procedures might place additional demands on optimal primer selection. One is cloning of the PCR product. The most commonly used methods are listed in Table 7.2; for a new sequencing-independent method see Chapter 9 by Guo Jun Chen. Restriction cloning requires the presence of a restriction enzyme recognition sequence, not too close to the 5’ end. Generally, a “clamp” of 8 to 10 bp is more than sufficient to ensure efficient digestion. However, the New England Biolabs catalogue (http://www.neb.com/) contains a useful table showing that some enzymes are able to cut much closer to the 5’ end.

### TABLE 7.1
Relative Efficiency with Which Taq Polymerase Extends 3’ Mismatches at 200 \(\mu M\) dNTPs \(^{11}\)

<table>
<thead>
<tr>
<th>Template</th>
<th>T</th>
<th>C</th>
<th>G</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>(2 \times 10^{-3})</td>
<td>1</td>
<td>(2 \times 10^{-2})</td>
<td>(10^{-3} - 10^{-4})</td>
</tr>
<tr>
<td>C</td>
<td>(10^{-2})</td>
<td>(&lt;10^{-4})</td>
<td>(&lt;10^{-4})</td>
<td>(4 \times 10^{-2})</td>
</tr>
<tr>
<td>G</td>
<td>(8 \times 10^{-2})</td>
<td>(&lt;10^{-4})</td>
<td>(&lt;10^{-3})</td>
<td>(&lt;10^{-4})</td>
</tr>
<tr>
<td>A</td>
<td>(10^{-3} \times 10^{-3})</td>
<td>(10^{-2})</td>
<td>(&lt;10^{-4})</td>
<td>(2 \times 10^{-4})</td>
</tr>
</tbody>
</table>

**Note:** Values shown in bold highlight mismatches with the lowest amount of nonspecific extension at 800 \(\mu M\) dNTPs. \(^{12}\)

### TABLE 7.2
The Most Common PCR Cloning Strategies and Their Requirements for Primer Selection

<table>
<thead>
<tr>
<th>Extra Requirements</th>
<th>Proofreading Compatible</th>
<th>Directional Cloning</th>
<th>Cloning Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Restriction digest</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>High(^a)</td>
</tr>
<tr>
<td><strong>Heterostagger</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Medium</td>
</tr>
<tr>
<td><strong>TA</strong></td>
<td>Yes</td>
<td>No(^b)</td>
<td>High</td>
</tr>
<tr>
<td><strong>Blunt</strong></td>
<td>Yes</td>
<td>No</td>
<td>Low</td>
</tr>
</tbody>
</table>

\(^a\) Depends on the efficiency of the digest.

\(^b\) Possible by an additional incubation step with Taq.

www.taq.ir
Primer pairs used to clone a gene for protein expression should be as close as possible to the region of interest. Unless the protein is to be fused to another, the forward primer should end in the start codon (ATG), the reverse primer in the complement of a stop (TAA/TGA/TAG; TAA is preferred). It is often useful to change wherever possible G or C to A or T for the first few codons after the translation initiation site. This will help reduce secondary structure and therefore improve expression.

III. OPTIMIZATION OF REACTION MIXTURE AND TEMPERATURE CYCLE

For many applications, maximum yield (and sometimes even specificity) are not required. In these, satisfactory results will be produced with little more than good primer design, and sequential testing of a range of 3 to 5 annealing temperatures and a magnesium concentrations. However, where PCR technology is stretched to its limits (see Table 1 in this book’s introduction, “PCR Technology: An Introduction to Current Trends and Innovations”) more careful optimization can be necessary to improve yield, specificity, and reliability (avoiding “dropout”).

A. MATERIALS AND METHODS

1. Reagents and Instrumentation

I developed the PCR reactions used in the examples for typing transgenic mouse strains at the Edward Jenner Institute for Vaccine Research, Compton, U.K. Genomic DNA was extracted from mouse tail samples using proteinase K digestion, followed by phenol/chloroform extraction. Primers were designed to amplify fragments of several human (trans)genes (Table 7.3) and ordered from GIBCO BRL (www.invitrogen.com) as desalted oligonucleotides. Transgenes were present at 1 to 10 copies per mouse genome as determined by Southern hybridization (not shown).

### TABLE 7.3
Reactions Optimized by the Taguchi Methods

<table>
<thead>
<tr>
<th>Gene</th>
<th>PCR Product</th>
<th>GC-Content</th>
<th>Primer Sequences (5′ → 3′)</th>
<th>Calculated T&lt;sub&gt;m&lt;/sub&gt;①</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-B27</td>
<td>437 bp</td>
<td>71%</td>
<td>GCCGCGAGTCCCGAGAGA</td>
<td>59.0°C</td>
</tr>
<tr>
<td>(human HLA-class I)</td>
<td></td>
<td></td>
<td>CCACGTGCAAGCATACATA</td>
<td>59.5°C</td>
</tr>
<tr>
<td>GrbB</td>
<td>403 bp</td>
<td>55%</td>
<td>ACACCTGGTCTAGGGAATG</td>
<td>54.6°C</td>
</tr>
<tr>
<td>(human TCRβ)</td>
<td></td>
<td></td>
<td>GAGACGGCCACAGAAAAGTG</td>
<td>60.7°C</td>
</tr>
<tr>
<td>GrbA</td>
<td>785 bp</td>
<td>46%</td>
<td>cacccgcaCCCGGGACCTG</td>
<td>69.2°C (35.7°C)</td>
</tr>
<tr>
<td>(human TCRα)</td>
<td></td>
<td></td>
<td>tttgttacCGCGGAGTTCTAATCCCTC</td>
<td>69.5°C (56.5°C)</td>
</tr>
</tbody>
</table>

Note: T<sub>m</sub> values were calculated using the nearest-neighbor method as implemented in the program OLILO (1989 to 2002, Wojciech and Piotr Rychlik). Monovalent ion concentrations were set to 50mM monovalent ions, MgCl<sub>2</sub> to the optimized value described below. The GrbA primers were designed originally for amplifying the Grbα gene segment from the human TCRα locus and cloning. The murine sequences flanking the restriction sites of the transgenic DNA samples on which the optimization was performed are not known. The T<sub>m</sub> values in brackets apply to the 3′ ends of the GrbA primers which are guaranteed to match the sample DNA and thus represent the minimum values in the first amplification cycle.

① Small case letters indicate bases that are potentially mismatched with the target sequence flanking the integration site of the transgene.

www.taq.ir
Magnesium chloride was obtained from BDH (http://www.bdh.com/) and stored as 1M stock, and aliquots of 50mM “working stock” solutions. MgCl₂ solutions were vortexed after thawing and prior to use because a concentration gradient can form during freezing. The other reagents and their suppliers were: dNTPs (Promega; http://www.promega.com/), Taq-polymerase (Bioline; http://www.bioline.com/) and betaine monohydrate* (Sigma, Poole, U.K.; http://www.sigmaaldrich.com/). All reagents were stored at −20°C; betaine as 5M stock, working dilutions of betaine and MgCl₂ (described in the spreadsheets) as 500-μl aliquots.

As described in the text, 20 μl reactions were set up in 250-μl reaction thin-walled tubes. The reactions were overlaid with 50-μl light mineral oil (Sigma) to prevent evaporation (lower-grade preparations may contain nucleases while autoclaving or UV irradiation for long periods may cause the formation of reaction-inhibiting hydrocarbons14).

Tubes were kept on ice during setup, then transferred quickly to a preheated thermocycler (DNA Engine Tetrad PCT225 from MJ Research; http://www.mjr.com/), prenatened at 94°C for 4 min, followed by 25 to 35 PCR cycles consisting of 1-min intervals each at 94°C, a variable annealing temperature, and 72°C, followed by a final extension at 72°C. PCR products were separated by electrophoresis in 1 to 2% agarose (FMC; http://www.fmcbiopolymer.com/) gels, 0.5 x TBE buffer and stained for 30 min in an equal volume of 1 mg/ml ethidium bromide in 0.5 x TBE.

Gels were placed under UV light and scanned using the Alpha Imager 2200 Documentation and Analysis System (Alpha Innotech Corp., San Leandro, CA; http://www.alphainnotech.com/). I quantified peak areas in autodetect mode with a rubberband-type baseline. Further details and tips can be found in Box 1.

2. PCR Optimization Spreadsheets

I chose to concentrate on four of the most convenient and commonly employed parameters in PCR optimization: annealing temperature, and the concentrations of MgCl₂, primers and a dsDNA destabilizing additive.14 To facilitate setup and analysis of the experimental design, I developed two spreadsheet programs, “Taguchi” and “5 × 5 Array.” These can be downloaded from http://www.crcpress.com/e_products/downloads/download.asp?cat_no=1184. Taguchi” and “5 × 5” also serve as passwords for unlocking the workbooks. The spreadsheets provide largely self-explanatory guides for planning, preparation, analysis, and documentation of the results: Cells in red font can be changed, others are protected. Blue font indicates cells whose values need to be checked. Moving the cursor over a cell with a red triangle in the upper right corner will reveal a text box explaining the content of this cell. Both workbooks share a similar structure. Sheet No. 1 lists the optimization parameters. Annealing temperature, primer concentration, and magnesium ion concentration are the most critical parameters in most types of PCR reactions for optimization. In addition, dsDNA destabilizing reagents are sometimes necessary to amplify templates with “difficult” sequence motifs. Since the latter are not well defined (“GC-rich," hairpins, etc.) including dsDNA destabilizing reagents in the first round of optimization is commendable. The spreadsheets can be modified for optimization of other parameters such as ramping rate or Taq polymerase concentration. The initial ranges for the tested parameters should be chosen wide enough so that points on both sides of the optimum are likely to be included. The values shown in Box 2 proved useful for all PCR systems studied here.

Sheet 2 provides pipetting schemes for the setup. I employed master mixes wherever possible to save time and minimize the effect of pipetting errors. Additional useful precautions are listed in Box 1. Also on sheet 2 are fields to enter “scores” for PCR results. Generally, these will be yield and/or specificity, but other readouts could be scored, as well.

* Both the monohydrate and anhydrous form are suitable, but betaine HCl solution is extremely acidic and will inhibit PCR.
### Box 1

**General precautions for PCR optimization with the Taguchi method**

- Run the initial experiment with both a negative and a positive control to avoid “optimizing” a nonspecific band of similar size. For this the volumes in the master mix should be doubled, except for the DNA which will be added after the master mix has been divided into two aliquots.
- It is critical to perform optimization under exactly the same conditions as the later routine application, i.e.,
  - PCR reagents (NH₄SO₄, NH₄Cl, or KCl Tris or Tricine, etc.)
  - DNA (plasmid, cDNA, or genomic; different complexity may require different stringency for priming)
  - Tube thickness and shape (some fit the heat block better than others)
  - Reaction volume (affects ramp rates)
  - Setup conditions (hot or cold start, setting up on ice or at room temperature)
  - Mineral oil overlay or heated lid (may affect ramp rates)
- Planning, setup, and PCR
  - Minimize pipetting steps using master mixes and (serial) working dilutions
  - Use volumes within the most accurate range of a given pipette type
  - Avoid pipetting small volumes of viscous solutions (use working dilutions or add to master mix where possible)
  - Mix very well (especially when adding DNA to master mix)
- Run all samples at the same time, preferably close to the center of the heat block
- Gel staining
  - Use equal concentrations of ethidium bromide in gel and running buffer. If gels are to be stained after the run, best equilibrate overnight before scanning.

---

### Box 2

**Recommended initial parameter ranges**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annealing temperature</td>
<td>$T_m$ as calculated by nearest neighbor method (50 mM monovalent salt, 2mM MgCl₂), values 5°C above and below.</td>
</tr>
<tr>
<td>Primer concentration</td>
<td>0.2, 0.6, and 1.0 uM for most PCRs; may vary depending on tendency to form primer dimers and number of potential mispriming sites in the template.</td>
</tr>
<tr>
<td>MgCl₂ concentration</td>
<td>0.5, 1.5, and 2.5 mM</td>
</tr>
<tr>
<td>Betaine concentration</td>
<td>0, 0.6, and 1.2 mM</td>
</tr>
</tbody>
</table>

Finally, sheet 3 shows graphs from which the optimum range for each parameter can be obtained. In case of the Taguchi method, scores from all tubes sharing a particular parameter value are transformed into “signal-to-noise levels,” which is somewhat less sensitive to outliers, especially dropout.¹² For the $5 \times 5$ arrays, values are simply normalized to a percentage of the sum of all scores. The formulas used in these calculations can be seen in the spreadsheets by selecting the appropriate cells.
3. Quantifying PCR Yield

Although the Taguchi method is designed to cope with outliers, reliable quantification of PCR yield is essential for its success. Scanning of ethidium bromide (EtBr) stained gels produced results equal or even better than hybridization with a $^{32}$P-labeled oligonucleotide probe (not shown). In any case, the detection method which seems most relevant to the later application should be chosen right from the start.

I ran samples on 1 to 2% agarose gels, 0.5 x TBE, followed by at least 30 min shaking and staining with EtBr in 0.5 x TBE (solution can be reused several times but will fade). Quantitation was done using “autodetect” mode and “rubberband” baseline option. Peaks that at first are not picked by the software are often detected after switching peak detection parameters back and forth. Alternatively, the program allows selecting peaks by hand.

Although ethidium bromide (EtBr) fluorescence is a convenient and relatively accurate readout for DNA quantity, two points require attention: Firstly, EtBr migrates in the opposite direction from DNA during electrophoresis. When EtBr is added to the gel only, bands lower in the gel will stain less strongly once the dye front has passed. For accurate quantitation using EtBr, the dye must be included in the running buffer at the same concentration or the gel stained after the run. Secondly, the fluorescence signal obtained by this method is affected by the sharpness of the bands in the gel. This is because DNA close to the gel surface will get stained more quickly than inside. When using gel combs of a different thickness or a PCR additive that affects band sharpness it is better to leave the gel in the fridge, wrapped in cling film, overnight and scan again the next day.

B. Examples and Results

1. Step 1: Taguchi Method

It was important that bands were detectable and scored significantly over background values (in the examples, peak areas >40) for at least 6 out of the 9 reactions. If more than 3 reactions failed, deduction of signal-to-noise level (SNL) optima became unreliable and the experiment needed to be repeated with one or more of the following changes:

- Wider spacing of values
- Higher number of PCR cycles
- Higher copy number of target molecules and/or better quality DNA

Where still no specific product is detectable, it is probably best to design a new primer pair. If at least 6 reactions resulted in a detectable product, optimum ranges could be estimated. Figure 7.1A and Figure 7.2A show the first step in the optimization of HLA-B27 and GrbB PCR, respectively. Although not strictly necessary, a second, and sometimes third, experiment was performed using previous estimates to narrow down the optima and check the consistency of the method. It is important to bear in mind, however, that at this stage only qualitative comparisons are possible between two different experiments (i.e., more or less yield before and after parameter change). Interestingly, the amplification of all tested genes showed consistent optima for MgCl$_2$ and annealing temperature but often V-shaped graphs for primer and betaine concentrations (Figure 7.1A).

2. Step 2: 5 x 5 Array

These results suggested that, at this stage, the Taguchi method was too insensitive to further improve PCR performance. V-shaped graphs could indicate “dropout” of a reaction with the middle value, or the presence of two local optima. These possibilities required a more quantitative
approach, and testing of all possible combinations of parameter values. Since the search had been narrowed down to two parameters, I performed a two-way titration as shown in Figure 7.1B. The program “5 × 5 Array” helps to set up and analyze such an experiment. After entering the scores, sheet 2 produces a contour graph similar to the one in Figure 7.1C, Figure 7.2B, and Figure 7.3A. The example data on the spreadsheet are from the experiment shown in Figure 7.3B. To illustrate the local optima, the two graphs have been merged, using experiments.
performed with the same master mix and a replicated series of primer dilutions at 0.3 M betaine to normalize the scores. Table 7.4 summarizes the new optima obtained with Taguchi and 5×5 Array optimization.

C. Discussion

Annealing temperature and magnesium chloride concentration, the parameters that are most commonly varied for PCR specificity and yield, also emerged as critical factors in the Taguchi protocol. So far, dsDNA destabilizing reagents have been only recommended for “GC-rich” targets, such as HLA-B27. However, the performance of all three PCR assays could be further improved by increasing the concentrations of both primers and betaine. Interestingly, at least the GrbA reaction showed clear evidence for two optima, one at low primer and betaine concentrations, and another in presence of unusually high amounts of primer and additive. It is unlikely that this second optimum would have been discovered by a conventional optimization strategy. The likely reason is the paradoxical effect of dsDNA destabilizing additives on PCR: by reducing the T_m of the template strands they facilitate primer annealing and extension but compromise both by reducing the T_m of the primers. The latter effect can be compensated by simultaneously increasing primer concentrations. Such compensation should be preferable to increasing MgCl_2 which, as a salt, might also stabilize dsDNA template (the combined effect of various buffer components is not straightforward). Importantly, high amounts of primers did not compromise specificity as long as the betaine concentration was raised accordingly (see FIGURE 7.2 Optimization of GrbB amplification. (A) As for HLA-B27 (Figure 7.1), two rounds of optimization with the Taguchi design resulted in consistent values for annealing temperature and MgCl_2 (56°C and 1.5 mM, respectively). (B) The 5×5 array allowed the unequivocal determination of optimal primer and betaine concentrations, however, no evidence for two optima was found for this PCR product. The possibility of a second maximum at >1.2 M betaine and >1.2 mM primer concentration (Figure 7.2A, first and second Taguchi optimization, respectively) was not tested.
Further experiments will show whether these are general rules for the PCR.

Apart from optimizing for yield and specificity, the strategy described here could, with the appropriate read outs, also be used for optimizing other characteristics such as DNA polymerase fidelity, *in vitro* mutagenesis, or even for applications other than PCR. Finally, it should be useful

---

**TABLE 7.4**

<table>
<thead>
<tr>
<th>PCR Assay</th>
<th>$T_m$ (°C)</th>
<th>MgCl$_2$ (mM)</th>
<th>Betaine (M)</th>
<th>Primer (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-B27</td>
<td>56</td>
<td>1.5</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>GrbB</td>
<td>56</td>
<td>1.5</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>GrbA</td>
<td>56</td>
<td>1.5</td>
<td>0.8</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Figure 7.1B for an example). Further experiments will show whether these are general rules for the PCR.
for developing a standardized protocol for comparing the efficiency of different PCR enhancers. (Due to the interdependence of dsDNA stability affecting variables, a novel additive may not necessarily work best at conditions optimized for another reagent.)

ACKNOWLEDGMENTS

I would like to thank my colleagues at the Edward Jenner Institute for Vaccine Research, Compton, U.K., where most of the experimental work described here was done, and especially Dr. Helen Bodmer, group leader of the autoimmunity group, for her support.

REFERENCES

Labeling and Cloning
8

Economic Fluorescent Labeling of PCR Products for Microsatellite- and Single-Stranded Conformation Polymorphism (SSCP) Analysis

Markus Schuelke

CONTENTS

I. Introduction................................................................................................................. 79
II. Materials and Methods............................................................................................... 81
   A. DNA Templates ................................................................................................. 81
   B. Primers............................................................................................................... 81
      1. Universal Fluorescent Primers .................................................................. 81
      2. Primers for Microsatellite and SSCP Analysis ...................................... 81
   C. Multiplexplex Microsatellite Labeling.......................................................... 82
   D. SSCP Analysis................................................................................................. 83
III. Results.................................................................................................................... 83
   A. Microsatellite Analysis.................................................................................... 83
   B. SSCP Analysis................................................................................................. 84
IV. Discussion............................................................................................................. 85

References.................................................................................................................. 86

I. INTRODUCTION

Most of the high-throughput analyses of PCR fragments are done on automatic devices. These are generally gel or capillary electrophoresis units with various laser detection systems (see also Chapter 11 by John M. Butler and Peter M. Vallone). In order for the PCR fragments to be detected by laser-light, they have to carry fluorescent tags. Colors currently used for these tags comprise 6-carboxy-fluorescein (FAM), hexachloro-6-carboxy-fluorescein (HEX), 6-carboxy-X-rhodamine (ROX), and tetrachloro-6-carboxy-fluorescein (TET) with emissions in the $\lambda = 517$ to 607 nm range and Cy5 and Cy3 in the infrared $\lambda = 572$ to 664 nm range. Primers labeled with such tags can be ordered from various commercial companies. The attachment of the fluorescent tag, however, is expensive. Especially in larger genotyping projects, the costs for fluorescent primers may pile up considerably. Most of these primers are used for less than 100 samples at a time and are later stored away forever.

Another application for fluorescent labeling of PCR products is the Single-Stranded Conformation Polymorphism (SSCP) analysis on automatic capillary or gel sequencers. Here, the sense-
as well as the antisense-strands are labeled with different fluorescent colors (e.g., Blue = FAM and Green = HEX) and are run on a nondenaturating gel. The various conformers of the single strands can easily be detected by the “pure” colors while the double strands are a “mixture” of both original colors. Since two labeled primers are required per SSCP assay, the setup is even more expensive than in microsatellite analysis.

I have developed a method by which the fluorescent labeling of PCR products can be achieved for a fraction of the cost encountered previously, without increasing the pipetting work. Labeling is achieved by a 3-primer protocol for microsatellite and by a 4-primer protocol for SSCP analysis. A sequence-specific forward or reverse primer is fitted with an adapter sequence such as T7, SP6, M13(-21) at its 5’-end. For microsatellite analysis the second sequence-specific primer is left unchanged. In the case of SSCP analysis the second primer is fitted with a different adapter sequence. In the first rounds of PCR the gene-specific primers are built into the PCR product thus incorporating their adapter sequences. These adapter sequences are used as annealing sites for the “universal” fluorescent primers. In subsequent PCR rounds the “universal” primer with its fluorescent tag is incorporated into the PCR product which can now be detected by laser-light (Figure 8.1). The “universal” T7, SP6, and M13(-21) primers with a fluorescent tag at their 5’-end have to be synthesized only once and can be used in all subsequent labeling reactions.

The feasibility of this approach for microsatellite labeling is demonstrated by a multiplex reaction with four markers at the locus of the NDUFS1 gene at chromosome 11q13. I am using this mapping approach in patients with isolated mitochondrial complex I deficiency. So far, seven different genes have been described to cause this biochemical deficiency. To minimize the sequencing effort, I first single out potential candidate genes by segregation analysis of the pedigree. This can be achieved by analyzing two highly polymorphic microsatellite markers at each side of the gene of interest.
The application of the fluorescent labeling method for SSCP analysis is demonstrated by analyzing a 552G>A splice-site mutation in exon 3 of the TTPA gene (GenBank XM_034473) of an index patient and his parents. Mutations in this gene cause Ataxia with isolated Vitamin E Deficiency (AVED) (OMIM #277460).

II. MATERIALS AND METHODS

A. DNA TEMPLATES

Human genomic DNA is isolated from whole blood by salt extraction, after lysing erythrocytes, as described previously. DNA for PCR reactions is diluted with Tris/EDTA-buffer to 50 ng/µl.

B. PRIMERS

1. Universal Fluorescent Primers

At the beginning, two fluorescent “universal” primers (Table 8.1) have to be synthesized commercially in the 200-nmol range. To guarantee uniformity of length, the “universal” primers should be HPLC-purified. The annealing temperatures of the universal primers should be quite low, at around 52°C. For multiplex reactions I use two universal primers to be able to distinguish microsatellites that have overlapping length ranges. The first 20 PCR cycles are performed at an annealing temperature of around 60°C. It is assumed that under these conditions the PCR reaction mainly proceeds with the gene-specific primers. However, since the 5′-tagged primer has only 1/10 the concentration of its reverse primer, it will be used up after these 20 cycles. In the second part of the PCR reaction the annealing temperature is lowered to 52°C for another 20 PCR cycles. The universal fluorescent primer can now “take over” and hybridize to the 5′ tag of the original gene-specific primer and sustain the PCR reaction. If multiplexing is used in routine diagnostic procedures, sometimes the method has to be optimized in order to obtain peak sizes that are sufficiently uniform. If this cannot be achieved by simply increasing the number of PCR cycles, the amount of microsatellite-specific forward and reverse primer may be varied. The markers should be tested first in a single reaction. The various steps to optimize a multiplex reaction for microsatellite analysis and possible pitfalls are reviewed in Reference 4.

2. Primers for Microsatellite and SSCP Analysis

The primers for the microsatellite-specific markers (Table 8.2) were designed according to general principles. The annealing temperature (TM) should be calculated with the nearest-neighbor algorithm and should be uniform (i.e., 60 ±1°C) for the whole multiplex set. At the 5′ end of either forward or reverse primer the T7 or M13(-21) adapter sequence (see also Table 8.1) is appended. The adapter sequence is not included for calculation of the TM.

The primers for SSCP of exon 3 of the TTPA gene (Table 8.2) anneal in the intronic region ca. 30 bp away from the exon–intron boundary. The forward primer is fitted with a T7 and the reverse primer with a M13(-21) sequence at its 5′ end.

<table>
<thead>
<tr>
<th>Universal Sequence</th>
<th>Color</th>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>T7 promoter</td>
<td>FAM</td>
<td>FAM-TAATACGACTCATATAGGG-3′</td>
</tr>
<tr>
<td>M13(-21) promoter</td>
<td>HEX</td>
<td>HEX-TGTAAACGACGAGCCATGT-3′</td>
</tr>
</tbody>
</table>
1. Prepare the sample mix:
   Sterile distilled water 12.0 µl
   10 x *Taq* DNA polymerase buffer 5.0 µl
   MgCl₂ [25 mM] 5.0 µl
   Universal FAM-T7 primer [10 µM] 2.0 µl
   Universal HEX-M13(-21) primer [10 µM] 2.0 µl
   Forward primer #1 [10 µM] 2.0 µl
   Forward primer #2 [10 µM] 2.0 µl
   Forward primer #3 [10 µM] 2.0 µl
   Forward primer #4 [10 µM] 2.0 µl
   T7 Reverse primer #1 [1 µM] 2.0 µl
   M13(-21) Reverse primer #1 [1 µM] 2.0 µl
   T7 Reverse primer #3 [1 µM] 2.0 µl
   M13(-21) Reverse primer #3 [1 µM] 2.0 µl
   dNTPs [10 mM each] 2.0 µl

   The above-mentioned quantities are for one tube. They have to be multiplied by
   the numbers of samples plus one extra for every 10 samples to allow for minor pipetting
   errors. Pipette 1 µl of template DNA [50 ng/µl] into 200 µl Eppendorf reaction strips and
   add 44 µl of the reaction mix. Care has to be taken that the concentration of the T7 and
   M13(-21) reverse primers are only 1 µM.

2. Prepare the *Taq* DNA polymerase mix
   Sterile distilled water 4.4 µl
   BSA 20 mg/ml 0.4 µl
   *Taq* DNA polymerase 5 U/µl 0.2 µl

   The above-mentioned quantities are for one tube. They have to be multiplied by
   the numbers of samples plus one extra for every 10 samples to allow for minor pipetting
   errors. The *Taq* DNA polymerase mix can be included in the sample mix if “hotstart”
   *Taq* DNA polymerase is used. If “standard” *Taq* DNA polymerase is used, heat the filled
   reaction vessels to 96°C and add 5 µl of the *Taq* mix after 5 min denaturation. This
   prevents the generation of unspecific amplifiers.
Economic Fluorescent Labeling of PCR Products

3. Program the thermocycler:
   96°C (5 min) initial denaturation
   20 cycles: 96°C (30 sec) 60°C (30 sec) 72°C (60 sec)
   20 cycles: 96°C (30 sec) 52°C (30 sec) 72°C (60 sec)
   72°C (10 min) final elongation

4. Analyze the PCR products
   2 µl of the multiplex reactions
   0.25 µl of a GeneScan Standard (e.g., ROX 500, Applied Biosystems)
   8 µl deionized formamide
   Denature the mixture at 96°C for 5 min, then keep on ice until loading. Run on an Applied Biosystems Capillary Genetic Analyzer 3100 with the filter set D and POP6 matrix for 30 to 45 minutes. Analyze the output with the GeneScan2.1 software (Applied Biosystems).

D. SSCP Analysis

1. Prepare the sample-mix:
   Sterile distilled water 20.0 µl
   10 x Taq DNA polymerase buffer 5.0 µl
   MgCl₂ [25 mM] 5.0 µl
   Universal FAM-T7-primer [10 µM] 2.0 µl
   Universal HEX-M13(-21)-primer [10 µM] 2.0 µl
   T7-Forward primer [1 µM] 4.0 µl
   M13(-21)-Reverse primer [1 µM] 4.0 µl
   dNTPs [10 mM each] 2.0 µl
   The above-mentioned quantities are for one tube. They have to be multiplied by the numbers of samples plus one extra for every 10 samples to allow for minor pipetting errors. Pipette 1 µl of template DNA [50ng/µl] into 200 µl Eppendorf reaction strips and add 44 µl of the reaction mix. Care has to be taken that the concentration of the T7 and M13(-21) Forward and Reverse primers are only 1 µM.

2. Prepare the Taq DNA polymerase mix as in Section II.C.2

3. Program the thermocycler as in Section II.C.3

4. SSCP analysis of PCR products
   2 µl of the PCR product
   0.25 µl of the GeneScan Standard (ROX 500, Applied Biosystems)
   6 µl loading buffer (deionized formamide, 10 mM NaOH, dextran/loading dye)
   Denature at 96°C for 5 min, and then keep on ice until loading. Run 2 µl of this mixture on an Applied Biosystems 377 Genetic Analyzer in a nondenaturating gel (0.5 x MDE, 10% Glycerol, 1 x TBE) at 16°C with 12 cm glass plates, filter set D, 2000 V, 40 mA, 20 W for 4 h with 2400 scans/h. Analyze the output with the GeneScan2.1 software (Applied Biosystems). Aberrant conformers in the case of heterozygous or homozygous mutant strands can thus easily be detected.8

III. RESULTS

A. Microsatellite Analysis

The results of the multiplex microsatellite analysis are depicted in Figure 8.2. The stutter peaks, characteristic of dinucleotide repeats, can be clearly seen. The marker fragments are aligned according to their lengths by the GeneScan2.1 software. All four microsatellite markers are informative at the NDUFV1 locus at chromosome 11q13. Since the marker lengths of D11S1889 and D11S4207 are overlapping or at close range at around 220 bp and the lengths of D11S1314 and
D11S4178 at around 380 bp, I chose to label them with two different fluorescent colors. Thus, I can differentiate between two markers even if they overlap.

### B. SSCP Analysis

The results of the SSCP analysis are depicted in Figure 8.3. Since the running patterns of the single strands do not primarily depend on their sizes, the gel readouts are aligned by the scan number of...
the laser scanner. In fluorescent SSCP the bands of the forward and of the reverse strand carry different fluorescent colors and can thus be differentiated. In the case of the 552G>A mutation, the two wild-type single strands run at around 6100 and 7700 scans and the mutant single strands at around 6800 and 7100 scans. A small portion of the single strands reanneals despite the presence of formamide. These double strands run faster than the single strands and are detected at around 5000 scans with overlapping blue and green colors.

IV. DISCUSSION

The protocols for universal fluorescent labeling are easy to perform and do not cause more pipetting work. They can be automated and used for high- as well as for low-throughput applications. Since the cost of the setup of a single fluorescent microsatellite marker is around $100 to $150, and of an SSCP analysis around $200 to $250, a lot of research funds can be saved by a one-time synthesis of two universal primers. The saving may be substantial in fine-mapping projects in which numerous
“rare” markers unavailable as commercial standard panels have to be run on a small number of samples.\(^9\) The method has been successfully employed in fields where no ready-made markers are available, such as in zoology\(^10\) and agricultural research.\(^11\)

It also offers the opportunity to incorporate a variety of different tags into a DNA strand by simply labeling a single universal primer. A variety of 5′-terminal modifications such as biotin or alkaline phosphatase or a 5′-end amidation for coupling DNA with peptides can thus be introduced into a DNA molecule.

In the present example the SSCP analysis is run on an ABI 377 Sequencer with a non-denaturing gel system. Capillary sequencers such as ABI310 or ABI3100 can also be used. In this case, however, the denaturing POP6 capillary matrix has to be replaced by a non-denaturing matrix (e.g., GeneScan polymer without TSR, Applied Biosystems).

REFERENCES

Universal Restriction Site-Free Cloning Method Using Chimeric Primers

Guo Jun Chen

CONTENTS

I. Introduction................................................................................................................. 87
II. Materials and Methods.............................................................................................. 88
   A. Enzymes and Reagents......................................................................................... 88
   B. Cloning Protocol.................................................................................................... 88
      1. Principle........................................................................................................... 88
      2. Primer Design.................................................................................................... 89
      3. Preparation of Insert and Linear Vector........................................................... 90
      4. Generation of 3'-Overhangs............................................................................. 91
      5. Phosphorylation and Ligation........................................................................... 91
      6. Transformation and Selection......................................................................... 91
III. Discussion.............................................................................................................. 92
References..................................................................................................................... 92

I. INTRODUCTION

Cloning of PCR products in order to express native proteins, to create chimeric genes, or to study interactions between DNA and proteins requires a method that allows the insertion of the amplified DNA fragment into a vector in a defined position and orientation. Neither the sequence of the DNA fragment nor that of the vector should be changed by the procedure. Routinely used methods such as restriction enzyme and T-A cloning require specific DNA sequences and are, therefore, not always suitable for the above tasks. Although several other methods have been developed to circumvent this problem (e.g., enzyme-free cloning, heterostagger cloning, auto-sticky PCR cloning, and restriction site-free cloning), their drawbacks may hamper them from being used widely.

A more recently developed method, “universal restriction site-free cloning,” can be used for most molecular cloning applications because it is independent of the DNA sequence of the insert and the vector. This cloning method (Figure 9.1) is somewhat similar to that of the sticky-end cloning technique using restriction enzymes. Instead of having a restriction site, however, primers contain a ribonucleotide that serves as a specific cleavage site for a rare-earth metal ion. Incubation of PCR products with the rare-earth metal ion introduces strand breaks, creating overhangs for cloning.
II. MATERIALS AND METHODS

A. ENZYMES AND REAGENTS

Expand High Fidelity PCR System (Roche Molecular Biochemicals)
T4 DNA ligase (Roche Molecular Biochemicals)
T4 polynucleotide kinase (New England Biolabs)
Deoxynucleotide triphosphates (dNTP) (Roche Molecular Biochemicals)
Lanthanum chloride (Sigma) or lutetium chloride (Aldrich Chemical)
Chimeric primers (synthesized by Microsynth GmbH)
SOC medium

B. CLONING PROTOCOL

1. Principle

The method uses two pairs of chimeric primers, each containing a ribonucleotide (Figure 9.1). One pair of primers is used to amplify an inserted DNA fragment and another pair is used to

FIGURE 9.1 Schematic representation of the universal restriction site-free cloning method. In the figure, the vector is in grey, while the insert is in black. Ribonucleotides are represented by the symbol ↑.
prepare a linear vector. The ribonucleotide, together with its 5’ linked DNA portion, can be removed by the treatment of the rare-earth metal ion such as La\(^{3+}\) or Lu\(^{3+}\) to break the phosphodiester bond between 3’-ribonucleotide and 5’-deoxyribonucleotide (Figure 9.2). Thus, the blunt-ended PCR products are converted into double-stranded DNA fragments with 3’ overhangs at both ends.

The location of the ribonucleotide in the primers governs the length of the sticky-end to be created. The role of the rare-earth metal ion treatment is analogous to a restriction enzyme digestion but there is no requirement for a specific DNA sequence. By designing the primers so that the 5’-end sequence in the primer for the insert is complementary to that for the vector, complementary 3’ overhangs between the insert and the vector will be generated during the rare-earth metal ion treatment. Thus, after phosphorylation of the insert and the vector to create T4 DNA ligase ligatable 5’ phosphate ends, a seamless new plasmid will be formed in the ligation step.

2. Primer Design

The primers used in this cloning method are to link two unrelated DNA fragments together at defined positions to form a new plasmid. In order to create complementary 3’ overhangs between the insert and the linear vector, draw a resulting plasmid map by inserting the target DNA fragment into the desired position in the vector. At the junctions between the two DNA molecules, design two pairs of primers of at least 20 bases in length for the amplification of the insert and the vector. The selected sequences should give a predicted melting temperature (T\(_m\)) of around 50 to 60°C during PCR. The following formula can be used as a thumb-rule to estimate the primer T\(_m\): T\(_m\) (°C) = 2 \times [total number of A and T] + 4 \times [total number of G and C].

The 5’-end sequence of each primer for the insert should be complementary to that of its corresponding one for the linear vector. One of the deoxyribonucleotides in the primer is replaced by a ribonucleotide with the corresponding base. The location of the ribonucleotide is determined
by the required length of the complementary overhang. In order to create an overhang of N bases, the Nth deoxyribonucleotide, numbered from the 5' end of the primer, should be replaced by its corresponding ribonucleotide.

Table 9.1 lists the two pairs of the chimeric primers that were used to generate PCR products with 10-base long 3' overhangs and that have successfully been used for cloning wild-type *E. coli* peptidyl-tRNA hydrolase to pDS56/RBSII.5

### TABLE 9.1
Chimeric Primers for the Construction of an *E. coli* Pth Expression Plasmid

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primer 1</td>
<td>CGTCATAGTUAATTTCTCCTTTAATGAA</td>
</tr>
<tr>
<td>Primer 2</td>
<td>GCAATAAGGCAGTTATGTGGCCTTAAAC</td>
</tr>
<tr>
<td>Primer 3</td>
<td>AACTATGACGATTTAAATTGTGTCCGCTT</td>
</tr>
<tr>
<td>Primer 4</td>
<td>GCCTTATTGGCTTTAAGGCCTGGAAT</td>
</tr>
</tbody>
</table>

*Note:* Primers are in 5’ to 3’ direction; nucleotides shown in bold and underlined are ribonucleotides.

3. **Preparation of Insert and Linear Vector**

1. Prepare two separate 100 µl PCR reaction mixture in a 0.5 ml Eppendorf tube (on ice), one for the vector and one for the insert:

   1X PCR reaction buffer with Mg²⁺ (supplied with the Expand High Fidelity PCR System)
   0.5 µg *E. coli* BL21 strain chromosomal DNA (insert) or
   5 ng pDS56/RBSII plasmid DNA (vector)
   0.5 µM of each primer (vector: 1 and 2, insert: 3 and 4)
   200 µM dNTPs each
   2.5 U Expand High Fidelity DNA polymerase (add last)
   Fresh redistilled H₂O to a final volume of 100 µl (add first)

2. Mix thoroughly and spin briefly to bring all solution to the bottom of the tubes.

3. Run the following three-step PCR programs on a thermal cycler:

   35 cycles of
   - 45 sec at 94°C
   - 1 min at 50°C
   - 1 min (insert) or 3 min (vector) at 68°C

   The final extension is 15 min at 68°C, then hold at 4°C.

4. Place the tubes in a thermal cycler and start the PCR programs. Be sure that the top heating cover is on. If the thermal cycler does not have one, add 50 µl mineral oil on the top of the reaction mixture to prevent evaporation.

5. Mix PCR products with 10 µl 10X DNA loading buffer and load on a 1% agarose TAE gel. Connect the electrodes of the gel box to the power supply in correct orientation and carry out electrophoresis at 5 V/cm in 1X TAE buffer containing 0.5 µg/ml ethidium bromide until the PCR products achieve good separation. Note that ethidium bromide is a potent carcinogen and must be handled with caution!

6. Excise the required DNA band from the gel under UV illumination as quickly as possible because UV light damages DNA (wear safety glasses and/or a protective shield: UV light also causes DNA mutagenesis and is harmful to the naked eyes).

7. Extract the DNA fragment from the gel slice using QIAquick Gel Extraction Kit (Qiagen) and resuspend in H₂O.
4. **Generation of 3’ Overhangs**

1. Mix 10 μl of purified PCR products of the insert or the linear vector with 90 μl freshly prepared 10 mM LaCl₃-NaOH (or LuCl₃-NaOH) pH 9.5 solution in 1.5-ml Eppendorf tubes. The activity of both Lu³⁺ and La³⁺ is pH dependent. Lu³⁺ has slightly lower activity than La³⁺ at pH 9.5, but it reaches its maximum activity at pH 8.0.³,⁸
2. Incubate the mixtures at room temperature or 50°C for 30 to 120 min.
3. Precipitate both DNA samples by using the ethanol precipitation method: Add 10 μl 3 M sodium acetate, pH 5.2, and 250 μl ethanol. Mix thoroughly and spin at top speed to bring DNA down.
4. Wash the DNA pellets three times with 70% ethanol and resuspend it in H₂O.

5. **Phosphorylation and Ligation**

1. Quantify both the vector and the insert (ethanol-precipitated La³⁺ or Lu³⁺-treated PCR products) on an agarose gel. The quantity of DNA can be estimated by running a small portion of the vector and the insert on an agarose gel, together with a series of known concentration of the DNA marker. Stain the gel with ethidium bromide and compare fluorescence intensities to calculate the quantity.
2. Incubate the insert and the vector at 37°C in 10 μl 1x T4 DNA ligation buffer using 10 units of T4 DNA polynucleotide kinase for 30 to 60 min.
3. Mix 10 to 20 ng of the T4 DNA polynucleotide kinase treated insert with 25 to 50 ng of the linear vector in an Eppendorf tube.
4. Add 1x T4 DNA ligation buffer to a final volume of 20 μl.
5. Heat the mixture to 60°C and keep this temperature for 2 min. Cool down slowly to room temperature. This step is not necessary for short overhangs.
6. Add 5 U T4 DNA ligase and mix thoroughly.
7. Ligate at room temperature for 2 h or 4°C overnight.
8. Clean the ligated DNA sample by phenol extraction and ethanol precipitation.

The last step can be omitted if not all of the DNA is used for transformation.

6. **Transformation and Selection**

1. Add 50 μl ice-cold *E. coli* Sure 2 electroporation competent cells to a 1.5 ml Eppendorf tube on ice.
2. Add the entire ligated DNA into the tube and gently but thoroughly mix with the cells.
3. Transfer the cell suspension into an ice-cold BioRad 0.2 cm electroporation cuvette. Be sure the cell suspension goes to the bottom, contacts both electrodes of the cuvette, and is free from air bubbles.
4. Put the cuvette in a BioRad Gene Pulser™ chamber, or an equivalent electroporation system, and perform electroporation with settings of voltage 2500 V, resistance 200 Ω, capacitance 25 μF. A time constant of ~5 msec is usually obtained.
5. Immediately add 450 μl of SOC medium, prewarmed to room temperature, to the electroporation cuvette directly after electroporation and gently mix well. Cells tend to die exponentially after electroporation if no SOC medium is added.
6. Transfer the above mixture to a 1.5-ml Eppendorf tube and put into a waterbath at 37°C for 30 min.
7. Plate 100 μl of the transformation culture onto an LB agar plate containing 100 μg/ml ampicillin.
8. Incubate transformation plate upside down at 37°C overnight (12 to 16 h).
9. Check the plate and pick a single colony for further analysis.
III. DISCUSSION

Universal restriction site-free cloning method indicates the future direction for DNA molecular cloning. Although it is a simple cloning method, care should be taken for primer design and DNA polymerase selection. The cost of chimeric primer synthesis, which is currently expensive, could be reduced to a minimum if standard 3’ overhangs were used in the preparation of linear vectors. For example, initial codon ATG and terminal codon TAA could be used as standard 3’ overhangs for the construction of native protein expression plasmids, or suitable restriction sites could be found in the vector. Since primers used in this method contain a ribonucleotide, mixed DNA polymerases such as Taq and Pwo should be used in order to make a full-length, blunt-ended PCR product in a single step. Taq DNA polymerase can use ribonucleotides as a template but will add an extra nucleotide at the 3’ ends of PCR products while Pwo DNA polymerase is able to remove the extra nucleotide to create blunt ends. It may be helpful to introduce an end-polishing step after the PCR step to obtain a more blunt-ended PCR product.

In addition, this technique might be useful for complementing other applications such as enzyme-free cloning,1 ligation-independent cloning2 but without introducing additional nucleotide sequences, DNA cloning without restriction enzyme and ligase,3 directional Topo cloning,4 and in vitro mutagenesis and evolution (Chapter 36 by Sarah B. Faulkner et al.). It may also be applicable for autosticky PCR cloning5 by creating 5’ overhangs of the PCR product using DNA polymerases like Pfu or Vent which do not use ribonucleotides as a template. In summary, methods using chimeric primers to produce defined 3’ and 5’ overhangs DNA fragment can be expected to have a big impact on molecular cloning techniques.

REFERENCES

Section II

Analytical Applications
Detection of Nucleic Acid Variation
10 High-Throughput Genotyping and the PCR

Kalim U. Mir and Jiannis Ragoussis

CONTENTS

I. Introduction.................................................................................................................97
II. Platforms Currently Available........................................................................................98
   A. Separation Methods....................................................................................................98
   B. Homogeneous Methods..............................................................................................99
   C. Array Methods..........................................................................................................101
      1. DNA Chips............................................................................................................102
      2. Microbeads..........................................................................................................102
      3. Optical Fibers......................................................................................................103
III. Future Perspectives........................................................................................................103
   A. New Ways for Using PCR in Large-Scale Studies ......................................................103
      1. PCR in mTAS......................................................................................................103
      2. Immobilized PCR...............................................................................................104
      3. Generic PCR.......................................................................................................104
      4. DNA Colonies.....................................................................................................105
   B. Alternatives to PCR for Large-Scale Studies ...............................................................105
      1. Signal Amplification.............................................................................................105
      2. Single-Molecule Technology...............................................................................105
   C. Concluding Remarks..................................................................................................106
Acknowledgments ..................................................................................................................107
References ................................................................................................................................107

I. INTRODUCTION

The earliest of methods to be applied to the analysis of DNA sequence variation were renaturation studies, which were used for obtaining a global view of variation between species.\(^1\) Southern analysis\(^2\) and the Restriction Fragment Length Polymorphism (RFLP)\(^3\) enabled the analysis of specific polymorphisms and showed the power of molecular tests in genetics. The desire to explore the sequence variation underpinning genetic traits has brought a need to perform molecular tests on a large scale.\(^4\) Single nucleotide polymorphisms (SNPs) are the most abundant form of sequence variation. Where the number of SNPs to be analyzed is small, initial optimization of PCR and genotyping can be followed by high-throughput screening in large populations. However, the study of complex traits\(^5\) by association studies requires analysis of sequence variation in both a large number of individuals and a large number of loci; here undertaking a prior optimization effort is too large a task.

Estimates of the number of SNPs that need to be analyzed for effective association studies have ranged between \(5 \times 10^4\) and \(1 \times 10^6\) in case and control populations containing thousands of
individuals. Current technologies struggle to provide genotyping for such scales of study in a facile, cost-effective, and high-throughput way.

The process of analyzing sequence variation goes through stages of sample collection, DNA purification, amplification of loci, genotyping, and data management. Efficient integration of these stages is important. The choice of format for each stage is influenced by the format in other stages. The most difficult stage to integrate is the PCR, which is required to provide sufficient material and enrichment for the allele in question to be typed accurately. Typically, a separate PCR reaction needs to be performed at each locus and each may require some optimization. Although a number of locus specific reactions can be multiplexed in a single reaction tube, primers that can work together have to be found, and 10-plex is typically the limit. The genotyping reaction itself also requires careful design and optimization. SNP alleles can be discriminated by a number of molecular processes: allele-specific hybridization, primer extension, oligonucleotide ligation, and the invasive cleavage reaction. Allele-specific hybridization would be the most straightforward method, but studies using array hybridization, where the number of SNPs analyzed were too large for prior optimization, gave low signal-to-noise of perfect match to mismatch, and struggled to correctly call heterozygous loci.

The efficiency of a hybridization reaction is not so much determined by the thermal stability of the heteroduplex, which is easy to calculate, as the on-rate which is restricted by access to sequence by secondary structure and is very difficult to predict. When hybridization is linked with an enzymatic process, signal-to-noise can be significantly improved. Many of today’s techniques utilize primer extension and require little optimization if the assay is designed using off-the-shelf software.

It is convenient here to describe current methods in the context of three major platforms (Figure 10.1), although the lines of distinction are not necessarily clear and the implementation of one platform may utilize aspects of another. The first two platforms comprise separation and homogeneous methods. These are inherently nonparallel, in that only one or a few SNPs are analyzed in a single reaction volume. The third platform, arrays, can be massively parallel and a very large set of SNPs can be analyzed simultaneously in a single reaction volume.

This chapter reviews these genotyping platforms and discusses their integration with the PCR. Given the cost and difficulties involved, is there a future for PCR in large-scale studies, at least in its present incarnation? We consider this question and look at the alternatives that are emerging.

II. PLATFORMS CURRENTLY AVAILABLE

A. SEPARATION METHODS

Separation methods (Figure 10.1A) are those that differentiate between assay products of different sizes. Gel electrophoresis provides good size resolution. The slab gel format is expensive and cumbersome while Capillary Electrophoresis (CE) has proven more cost-effective and amenable to automation. Minisequencing assays performed on this platform are robust and require little optimization in singleplex mode. Denaturing High Performance Liquid Chromatography (DHPLC), which many laboratories use for SNP discovery, can also be used for genotyping but requires more optimization. In recent years, the introduction of Mass Spectrometry (MS), mainly Matrix-Assisted Laser Desorption Ionization (MALDI), has provided very high resolution by separating ions on the basis of mass/charge ratio. Although a serial process, MS offers high speed. Typically, an assay involves primer extension across the variable site. A great advantage is that if priming occurs from an incorrect sequence, then it can usually be detected by a mass that is different from the desired product. Current systems are partly homogeneous in that all reactions prior to MS are done in a single tube, with stepwise addition of reagents. Initial results have demonstrated that this method may be suitable for DNA pooling approaches enabling association studies at low cost. SNP allele frequency differences between DNA pools (300 to 1500 individual samples) can be determined (at a control allele frequency of 0.1 the method has a 95% power to detect allele frequency differences of 0.07), thus greatly reducing total genotyping costs.
High-Throughput Genotyping and the PCR

A fully integrated system is highlighted in the flowchart in Figure 10.2. The full costings for genotyping with this system are given in Table 10.1. It is likely that mass- and/or charge-coding strategies for high levels of multiplexing will be needed in the future. Protocols for an example of these techniques can be found in Chapter 12.

B. HOMOGENEOUS METHODS

Homogeneous assays are carried out entirely in one medium and require no separation of reactants from products (Figure 10.1B). Typically, all reaction components are placed in a single tube at the beginning of the assay, and fluorescence resulting from a molecular process is monitored. There must be one probing reaction per tube unless the assays within each tube are encoded by different fluorophores so that they can be distinguished. Advantages are direct readout and removal of post-assay processing. Reactions may be carried out in microtiter plates, proceeding directly in the same reaction wells as the PCR. Some homogeneous assays, such as the Amplification Refractory Mutation System (ARMS) and its derivatives have the important advantage that the assay involves amplification as part of the process, allowing direct analysis of genomic DNA. These methods rely on DNA strains that emit very little fluorescence until they bind to double-stranded DNA. However, there is no way of telling whether what has become labeled is the desired product.

FIGURE 10.1 The three major platforms for performing genotyping assays based on PCR. Examples of methods based on (A) separation of products, (B) homogeneous assays, and (C) methods based on arrays.
or something different. Another approach that uses DNA-binding dyes, Dynamic Allele Specific Hybridization (DASH),\textsuperscript{23} can be considered a descendant of the renaturation methods used more than three decades ago. Although DNA-binding dyes are retained, they are used here for monitoring the melting profile of the probe-target heteroduplex, rather than just to give a positive or negative readout. Variants of the technique will increase scale of analysis that can be performed.\textsuperscript{24}

A number of assays have been developed based on fluorescence resonance energy transfer (FRET) which measures the proximity of donor and acceptor dyes. The aim in many cases is for the acceptor to quench fluorescence.\textsuperscript{20,25} “Molecular beacons” contain a probe embedded in a stem loop secondary structure which becomes disrupted upon hybridization of the target and thereby de-couples donor from quencher.\textsuperscript{25} The beacon may be attached to a PCR primer.\textsuperscript{26} A positively charged conjugated polymer has been shown to electrostatically bind along a target DNA chain to amplify, by a FRET mechanism, the signal from a fluorescein molecule attached to a PNA probe.\textsuperscript{27} The donor-acceptor concept has also been extended to bead technology in a sensitive proximity assay.\textsuperscript{28} Another way of performing an assay without the need for separation of the product from

\begin{figure}
\centering
\includegraphics[width=\textwidth]{genotyping.png}
\caption{Genotyping using Sequenom’s MS system at the Wellcome Trust Centre for Human Genetics, U.K. The core group designs assays for the SNPs that include the PCR primers and conditions as well as the corresponding extension primers and stop mixes. Ordering of the primers is web based. The Centre’s research groups perform their PCR reaction on 384 well blocks provided by the Core group. The Core group provides sample sheets on the web that are filled in by the groups and made available to the Core through a server. Using this organization up to approximately 7000 reactions can be processed daily corresponding to between 20,000 to 50,000 genotypes according to the degree of multiplexing. The advantage of the Sequenom system is that a complete integrated package is provided. It should be noted that the process is not entirely under robotic. However, where user handling is required, the introduction of human error is minimized by barcoding plates and by having an effective tracking system based on electronically filled-in sample sheets. Suppliers of oligonucleotides need to have an efficient web based ordering and tracking system. The quality of the oligos is extremely important. This is followed up by asking for MS data or by checking the primers on the MS. A breakdown of the costs for all consumables is given in Table 10.1.}
\end{figure}
free fluorescent nucleotides is to detect the change in Fluorescence Polarization (FP) of dye-labeled primers when they are extended. Some assays are homogeneous in that only one medium is required for processing the samples as well as detecting results, but they are not closed tube assays in that additional components need to be added during the course of the reaction. One such technique, Pyrosequencing, typically generates around 50 bases of sequence data across the locus of interest and can access insertion–deletions (Indels) as well as SNPs (protocols can be found in Chapter 18 by Pål Nyrén).

C. ARRAY METHODS

The array platform immobilizes a repertoire of probes at the solid phase on which assay reactions are conducted simultaneously in a single reaction volume (Figure 10.1C). This includes arrays in which the repertoire is attached to a single planar surface such as the DNA “chip” arrays. It also includes arrays in which the repertoire is distributed over discrete solid-phase surfaces such as beads or fibers. The advantages of these methods are massive parallelism, which makes them amenable to the largest scale of studies, and because such a large number of reactions can be done in a single reaction vessel, there are orders of magnitude savings in the use of reagents and hence costs. Although reactants must be separated from products, it is done with ease because the product

---

**TABLE 10.1**

**A Breakdown of Consumable Costs for Running the Sequenom System Assays in the U.K.**

<table>
<thead>
<tr>
<th>Product</th>
<th>Supplier</th>
<th>Singleplex</th>
<th>4 plex</th>
<th>5 plex</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PCR Reaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primer set 0.04 OD</td>
<td>Metabion</td>
<td>0.6p/reaction</td>
<td>0.6p</td>
<td>0.6p</td>
</tr>
<tr>
<td><em>Taq</em> Polymerase</td>
<td>Qiagen</td>
<td>3.4p/reaction</td>
<td>0.9p</td>
<td>0.68p</td>
</tr>
<tr>
<td>AB gene 384 well plates</td>
<td>AB gene</td>
<td>0.8p/reaction</td>
<td>0.2p</td>
<td>0.16p</td>
</tr>
<tr>
<td>dNTPs</td>
<td>Stores</td>
<td>1.5p/reaction</td>
<td>0.33p</td>
<td>0.26p</td>
</tr>
<tr>
<td>Microseal “A” film</td>
<td>GRI</td>
<td>0.3p/reaction</td>
<td>0.08p</td>
<td>0.06p</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.4p/reaction</td>
<td>2.11p/reaction</td>
<td>1.76p/reaction</td>
</tr>
<tr>
<td><strong>Extension Reaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pipette Tips</td>
<td>Beckman Coulter</td>
<td>0.8p/reaction</td>
<td>0.2p</td>
<td>0.16p</td>
</tr>
<tr>
<td>Plate sealers</td>
<td>GRI</td>
<td>0.3p/reaction</td>
<td>0.08p</td>
<td>0.06p</td>
</tr>
<tr>
<td>Sarstedt 96 well “v” bottomed plates</td>
<td>Sarstedt</td>
<td>1.3p/reaction</td>
<td>0.33p</td>
<td>0.26p</td>
</tr>
<tr>
<td>SpectroCHIP — 384 elements</td>
<td>Sequenom</td>
<td>34p/reaction</td>
<td>8.6p</td>
<td>6.8p</td>
</tr>
<tr>
<td>Shrimp Alkaline Phosphatase</td>
<td>Sequenom</td>
<td>3.7p/reaction</td>
<td>0.9p</td>
<td>0.74p</td>
</tr>
<tr>
<td>MassEXTEND Enzyme</td>
<td>Sequenom</td>
<td>13.7p/reaction</td>
<td>3.4p</td>
<td>2.74p</td>
</tr>
<tr>
<td>HME Buffer</td>
<td>Sequenom</td>
<td>0.34p/reaction</td>
<td>0.1p</td>
<td>0.07p</td>
</tr>
<tr>
<td>Calibrant</td>
<td>Sequenom</td>
<td>0.1p/reaction</td>
<td>0.025p</td>
<td>0.02p</td>
</tr>
<tr>
<td>SpectroCLEAN</td>
<td>Sequenom</td>
<td>2.6p/reaction</td>
<td>0.66p</td>
<td>0.52p</td>
</tr>
<tr>
<td>Dideoxy nucleotide stop Mix</td>
<td>Sequenom</td>
<td>5.4p/reaction</td>
<td>1.35p</td>
<td>1.08p</td>
</tr>
<tr>
<td>Extension Primer 0.2OD</td>
<td>Metabion</td>
<td>0.5p/reaction</td>
<td>2p</td>
<td>2.5p</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60.94p/reaction</td>
<td>17.32p/ reaction</td>
<td>14.69p/reaction</td>
</tr>
<tr>
<td><strong>Total in GBP (£)</strong></td>
<td></td>
<td>67.34p/reaction</td>
<td>19.43p/ reaction</td>
<td>16.45p/reaction</td>
</tr>
<tr>
<td><strong>Total in U.S. Dollars ($)</strong></td>
<td></td>
<td>$1.01/reaction</td>
<td>$0.29/ reaction</td>
<td>$0.25/ reaction</td>
</tr>
</tbody>
</table>

Note: These include all plasticware and molecular biology consumables (primers are calculated for twenty 384 well plates). First column from left: Reagents, followed by Supplier used by the Core group, and then costs for singleplex, tetraplex, and fiveplex reactions. At a throughput of over one million genotypes/year consumable cost savings in the order of 50% are possible. The total cost of assays in dollars is also for reference but does not reflect the actual cost of these assays in the U.S. which are likely to be less than in the U.K.
signal becomes fixed on the solid phase and the reactants can simply be washed away. Array based methods can be homogeneous in the sense that reactants do not need to be separated from the products if, for example, beacons are attached to the solid phase.32,33

Arrays may carry probes which are specific for loci of interest or they may be “universal.” Universal arrays may be of two types. The first carry every sequence of a given length, “n-mers,” and may be used to anchor a SNP typing reaction.34,35 The second type of universal array carries a repertoire of sequences used to decode the results of an assay which is carried out in solution with probes bearing, in addition to the SNP-specific sequence, unique identity tags known as “bar-codes” or “zip-codes.”36 Bar-coded assays can be decoded by hybridization to an array which carries probes complementary to each bar code. The major advantage is flexibility, in that a single array can be used to analyze any target and new arrays do not need to be manufactured specifically for each project.

1. DNA Chips

High-resolution printing such as photolithography and ink-jet technologies, coupled with parallel synthesis methods, has led to the high density “chips” of today.31 The advantage of the DNA chip format over distributed array systems is that the identity of a probe is simply specified by the location on the planar surface, and no complicated coding system is needed to identify probes. Miniaturization increases the number of assays that can be performed and reduces reaction volumes, resulting in savings in reagent costs.33 Hybridization stringency can be controlled by producing electric fields in the region of the probes and the discrimination of mismatches can be enhanced by reversing the polarity of the field.38 Arrays can carry probes immobilized in the 5’ to 3’ orientation for primer extension by DNA polymerase,11,39 and this has been used to generate the first generation of human chromosome Linkage Disequilibrium (LD) maps40 (see Chapter 13 by Maido Remm et al.). Errors can be minimized by the use of the Apyrase enzyme, which discriminates against misincorporations due to their slower incorporation kinetics.41 An alternative approach is to perform PCR followed by tagged primer extension reactions in a plate format and then to identify the extension products by hybridization to oligonucleotide bar-code arrays42,43 (SNP code by Orchid Biosciences).

The microarray format allows analysis of a massive number of loci but is not so well disposed for analyzing a large number of samples. However, making hundreds of copies of the same array is feasible with the printing technologies that are in use today. Even a single array can be useful if pooled DNA is analyzed; a minority allele at <1% frequency has been detected in one study36 and at a 2 to 5% frequency in another.4 “Arrays” of arrays have been introduced in which each array is at a distinct location on a single surface and a manifold is used to apply different samples to each of them.11 If the “array of arrays” approaches leading to 5,000 to 10,000 genotypes per slide are combined with sample pooling (100 to 200 individual DNAs), then it would be possible to generate up to 1,000,000 genotypes per slide.4

If arrays are made by parallel in situ synthesis, the expense of making thousands of separate probes/primers in solution is eliminated. Until now, very large scale arrays could only be made by a few industrial manufacturing facilities but with recent introduction of a commercial array synthesizer (Febit, Germany) it will become entirely possible to make arrays in any laboratory by parallel in situ synthesis.

2. Microbeads

An alternative means for carrying out a large number of reactions simultaneously in a single reaction volume is to form discontinuous arrays in which the repertoire is distributed on microbeads or microspheres.5,6 When suspended, microbeads occupy three dimensions, and consequently their reaction kinetics are purportedly superior to DNA chips which occupy two. Also, because beads of very small sizes can be manufactured, the repertoire of probes that can be generated on beads
High-Throughput Genotyping and the PCR

is larger than DNA chips. But the fact that they can not be addressed spatially means that assays in which they are used are more difficult to implement in practical terms. Where the result of assays in DNA chip arrays can simply be decoded by analyzing the spatial location of a signal, the identity of beads has to be encoded in some way and then decoded upon analysis, or the identity of probe on a bead must be obtained by sequencing it. Currently, a flexible coding system is available in which microspheres are dyed with different ratios of fluorophores to generate a repertoire. Up to 100 “colors” are available when two fluorophores are combined at different ratios. More are possible if a third fluorophore is added. Semiconductor quantum dot nanoparticles, which are excited at a single wavelength but can emit at different wavelengths depending on their size, have been embedded at different ratios into microspheres for coding. In one strategy, Massively Parallel Signature Sequencing (MPSS), the probes on the beads, rather than being coded, are sequenced by a novel strategy. Beads are covered with sequences, which bind to complementary tagged amplicons derived from a single molecule. The beads are loaded into a two-dimensional chamber so that their position is fixed before a parallel process determines the sequence. It should be relatively straightforward to develop bead approaches so that they can be applied to samples prepared in the microtitre format.

3. Optical Fibers

The second type of discontinuous array is distributed on fiber-optic bundles, with each probe sequence on an individual fiber. Probes may be attached to the end of the fibers or along the outside of the fiber. The advantage of these types of array is that any sample solution into which the fibers can be dipped can be probed. An array of fiber bundles (arrays) can be created in which each bundle is used to analyze a different sample. Sequences can be placed on known fibers which are then assembled into arrays. Methods of combinatorial synthesis, either by directing light down fibers or by dipping the fibers in a split-and-mix scheme can also be imagined. A system that combines beads and fibers has also been described. Here, beads carrying probes are randomly self-assembled onto the ends of fibers. The identity of the beads on each fiber is determined by sequencing. The system can give high-throughput and is currently implemented in the bar-code format with reactions done in microtitre plates allowing approximately a thousand reactions on the individual samples in each well. The generation of a million genotypes per day is possible at a low cost with this system (data from Illumina, Inc.). This method has been commissioned to generate 15% of the data for the current international effort to construct the human SNP haplotype map (HapMap, http://www.genome.gov/page.cfm?pageID=10005338). At least one representative of each platform described in Section II is available commercially, suggesting that a minimum standard of reliability has been achieved. Which of these methods will offer the best ratio of robustness to throughput will be revealed within the next 3 years through the construction of “HapMap.”

III. FUTURE PERSPECTIVES

A. NEW WAYS FOR USING PCR IN LARGE-SCALE STUDIES

1. PCR in μTAS

The expense associated with PCR reactions suggests that it would be beneficial to perform analysis in small volumes with smaller reagent quantities. Efforts have been made to perform PCR in Micro Total Analytical Systems (μTAS) or “lab-on-a-chip,” including a continuous flow PCR performed in a network of zones, where each zone is kept at a different temperature. The length of channel in each temperature zone determines the length of each temperature cycle. In another approach, both forward and reverse PCR primers can be synthesized in a single same array microstructure element. Then each element can be isolated from the other elements in the array and the oligonucleotides can be cleaved from the surface to act as primers in PCR in solution.
(Xiaolian Gao; personal communication). This would be a good way of creating primers for large scale ARMS PCR. The primers could be distributed to separate cavities for PCR in a μTAS. These systems could significantly increase throughput by reducing the time it takes to do PCR from hours to minutes, due to the reduction or elimination of ramping times, better heat dissipation and the reduction of the cycling times themselves.

2. Immobilized PCR

Immobilized PCR has been performed by attaching one primer within an array gel pad while retaining the other in solution. In addition, both primers have been immobilized to a single element on a planar surface, causing the amplified product to form “bridges” (see Figure 10.3). Using the same principle, locus-specific amplification can be conducted on polyacrylamide microspheres. Attaching both primers to a single locality on a surface can facilitate multiplexing because, compared to a solution-phase PCR, primers from different pairs are prevented from cross-reacting with each other.

3. Generic PCR

Locus-specific PCR for even 50,000 SNPs will require $1 \times 10^5$ primers, a vast number even for commercial oligonucleotide suppliers and certainly difficult to handle effectively. Generic PCR purports to cut it down to just one primer by amplifying sample DNAs to which “generic” clamp sequences have been ligated. To reduce genome complexity, repetitive sequences were suppressed by cot-1 DNA and this was found concomitantly to reduce mismatch error. An alternative “generic” PCR method, degenerate oligonucleotide primer (DOP)-PCR, uses a single primer whose sequence is partly redundant and partly specific and acts as both forward and reverse primer to amplify a subset of fragments of the genome. Other single primers, carrying different specific sequences amplify other subsets. Then SNP typing is done on these reduced genomic complexity subsets. Multiple-strand displacement, an isothermal whole-genome amplification procedure, provides good representative amplification of the human genome. These emerging methods suggest that amplification without locus-specific primer sequences may be feasible. For example, one recently introduced commercial product, based on analysis on beads in fiber-optical arrays (Illumina

**FIGURE 10.3** DNA colonies. (A) Schematic of bridge amplification. A mixture of generic forward and a reverse primers is spread on a surface or within a gel, leaving a free 3' end. The target genome is fragmented and ligated at each end with sequences corresponding to these generic primers. In the first cycle, the target anneals to, for example, the forward primer and a complementary strand is synthesized. In the second cycle, the target is denatured from the immobilized complementary strand, the 3' end of which is able to anneal to the immobilized reverse primer and act as template for complementary strand synthesis. In the third cycle, both forward and reverse strands are denatured and then are able to anneal to unextended primers. Complementary strands to each are synthesized. (B) DNA colonies on a surface stained with SYBR Green I observed under epifluorescence microscopy. DNA colony size: ~1-2 μm. A density of over 10,000,000 colonies/cm² can be achieved (courtesy of Manteia Predictive Medicine).
High-Throughput Genotyping and the PCR

Inc.), uses generic PCR rather than locus specific amplification. Nevertheless, 50,000 SNPs would still require 50,000 oligonucleotides to type them, even after generic amplification. Thus the savings compared to ordinary ARMS PCR would be 50%.

4. DNA Colonies

As with clonal colonies of cells where a single cell seeds a colony, nucleic acid colonies can be seeded by single molecules. Perhaps one of the most exciting emerging uses of PCR in a large-scale context is PCR amplification of single DNA molecules to form immobilized DNA colonies. This has been demonstrated both on a planar surface with generic primers (see Figure 10.3) and within a gel (“polonies”) using either generic or locus-specific primers. Once single molecules from the genome are “cloned” and amplified in one of these ways, further analysis can be performed. Genotyping has been performed using allele-specific primers and haplotyping by multiple polony reactions on a single chromosomal fragment molecule. Even direct in situ sequencing is being attempted on generically amplified isolates derived from single molecules (Laurent Farinelli, personal communication; Robi Mitra). If a whole genome can be sequenced in this way there will be no more need for locus-specific amplification and genotyping.

B. Alternatives to PCR for Large-scale Studies

The field would be transformed if the need for preamplification of loci were eliminated. Particularly, massively parallel array-based analysis would allow ultrahigh throughput if it were not restricted by the need for PCR. Can locus-specific PCR be replaced?

1. Signal Amplification

Rolling Circle Amplification (RCA) generates multiple copies of a sequence by continuous polymerization round a circular template in an isothermal reaction. However, one commercial program to apply RCA on the large scale has been abandoned, which might suggest difficulties with the approach. Another method that offers signal amplification is the homogeneous invader assay, but in practice the use of preamplification tends to be retained. The signal from probes labeled with metallic nanoparticles can be amplified sufficiently by silver enhancement to enable detection by a flat-bed scanner. Furthermore, silver enhancement of metallic nanoparticles can be detected by Surface Enhanced Raman Scattering (SERS) and by electrical transduction, which further extend the detection limit. Demonstration of a facile way to incorporate the particles in assays is needed. The drawback of the signal amplification approach is that a small error will be amplified and this may ultimately rule it out as the method of choice in an area where accuracy is paramount.

2. Single-Molecule Technology

New techniques are being developed that forgo traditional “bulk” biochemical methods that analyze the average signal from an ensemble of molecules and instead examine single molecules. A single binding event or reaction can be amplified by RCA or by labeling with nanoparticles and a number of techniques have been developed that can start with a single molecule and then perform PCR amplification; these include MPSS, DNA Colonies, and Digital PCR. A commercial SNP typing system based on Fluorescence Correlation Spectroscopy (FCS) of multilabeled single molecules has recently been introduced (Olympus/Evotec, OA). However it is not a significant departure from other homogeneous techniques because even though single molecules are detected as they pass through the small focus volume of a laser, the assay strategy still retains the PCR step. The signal from a single molecule does not need amplification to be detected, as single fluorophore labels can emit enough photons to be detected if background noise is sufficiently suppressed. Coincident single-molecule detection of two PNA probes (each labeled with a single
fluorophore) as they hybridized to proximal sequences on genomic DNA (passing through a sheath flow) provided the specificity to detect an unamplified, single-copy target DNA molecule within a complex genomic background in a homogeneous assay. Single-fluorophore detection opens up the possibility of simple assays based on single molecule counting on surfaces or in solution. Other techniques analyze the DNA as a polymer either flowing through channels or on a surface. DNA polymers on a surface have been probed at SNP sites using tagged probes that can be detected by the AFM and by fluorescent probes.

If large amounts of data can be handled efficiently, sequencing would offer a number of advantages over typing SNPs. Currently, major efforts are made trying to sequence DNA by monitoring the incorporation or removal by enzymes of nucleotides that are base-specifically labeled. The most direct approach would be visualization of DNA bases at the nanoscale; for example, individual bases in single stranded portions of a DNA molecule have been imaged, using a scanning tunneling microscope (STM) (Hiroyuki Tanaka, personal communication). Sequencing methods are also being developed that involve passing single DNA polymers through biological or nanofabricated pores and making single-channel recordings of impedance changes due to the passing bases. However, such “near-field” approaches may not be able to compete with the high throughput that might be possible with several of the “wide-field” detection methods, which track fluorescent-labeled sequencing events on many thousands of molecules in parallel, that are also under development.

C. CONCLUDING REMARKS

As plans are afoot for massive studies of sequence diversity in populations (e.g., the U.K. BioBank) and very large-scale association studies are being advocated, there is a need for systems that integrate, on the large scale, all aspects of the experiments and data handling/analysis. Systems are emerging that perform most required tasks, and workflow software can help organize and track the whole system. However the cost of analysis (~ $0.35 per SNP in 4- to 5-plex reactions) is still too high for wide-scale application, and PCR is a major limiting factor in achieving high throughput. Technologies that circumvent PCR, such as those based on single molecule examination, would increase throughput and bring down costs but are faced with the formidable complexity of the human genome which impacts the specificity with which a desired locus can be targeted.

The consequences of incorrect base calls in genetic studies can be significant and it has been speculated that up to 5% error might be inherent in today’s techniques. Typing errors can originate from ambiguities in base-calling, cross contamination, and incorrect sample labeling. Errors from the latter two sources will minimize with automation and bar-coding. Also, pooling of sample DNA would minimize the representation of such errors. Errors should not be ignored and a means for verifying the reliability of a base call must be included into systems. Then, as long as low confidence base calls are identified, they can be eliminated or treated with caution. When very large-scale analysis is enabled, the capacity to tolerate missing data due to low confidence base calls or just reactions that have not worked will concomitantly increase. For example, one of the few studies of a high density of SNPs (on chromosome 21), which was performed by array resequencing technology without prior probe optimization, proved informative despite only around 65% of SNPs over this area being typed with confidence. This figure improves with the use of enzymatic extension (Asper, SNuPE) and is also likely to improve with the use of the enzyme apyrase.

In general, costs can be reduced by minimizing reaction volumes through miniaturization and by reducing the number of reactions that need to be performed by Multiplexing, DNA pooling, and the selection of SNPs that “tag” conserved haplotype “blocks” (regions of high LD). Array technology has the potential for further savings as already detailed in Section II C.

Today PCR remains a mainstay of genotyping technologies but emerging approaches could significantly change the manner in which it is used in large-scale studies or even eliminate it altogether. Methods for inexpensive and rapid sequencing of whole human genomes using single
molecule technology or DNA colonies are being explored. This would enable access to disease-causing mutations directly, including those to which the common SNPs do not associate through LD. It could eventually open up an era of personalized medicine in which health management is informed by an individual’s genome sequence.

ACKNOWLEDGMENTS

We would like to thank ABI, Anthony Brookes, Sequenom, Manteia Predictive Medicine, and Asper Biotech for permission to utilize their images in the figures. We would also like to thank Rob Mitra for sight of preprints, Xiaolian Gao for discussions, and Laurent Farinelli for comments on the manuscript.

REFERENCES

High-Throughput Genotyping and the PCR


High-Throughput Genetic Analysis through Multiplexed PCR and Multicapillary Electrophoresis

John M. Butler and Peter M. Vallone

I. INTRODUCTION

The measurement of DNA variation between individuals in the form of length or sequence polymorphisms plays an important role in a number of fields, including medical diagnostics, agricultural genomics, genetic mapping and linkage studies, and human identity testing and forensics. The growth in these fields and the demand for high-throughput typing information has generated considerable interest in the use of multicolor fluorescence technology. In addition, the introduction of multicapillary electrophoresis systems has improved automation and ease of use over previous gel electrophoresis approaches. The combination of multicolor fluorescence and multicapillary instrumentation was integral to the accelerated pace of the Human Genome Project in recent years and the earlier than expected completion of the first draft sequence of the human genome.

Methods involving the use of the polymerase chain reaction (PCR) combined with capillary electrophoresis detection have become wide spread. The combination of multiple capillaries in an array format greatly increases the potential throughput of a DNA analysis system.
commercially available systems, such as the 16-capillary array ABI PRISM® 3100 Genetic Analyzer, fluorescence from dye-labeled DNA molecules is displayed onto a charged-coupled device (CCD) as the DNA passes the detection region during electrophoresis (Figure 11.1). The ability to use both size and color for DNA fragment analysis has greatly expanded the possibilities for multiplexed DNA analysis.4–5

Short tandem repeats (STRs) are regions of DNA that vary in their length due to a 2 to 6 bp tandemly repeated sequence that may occur 5 to 50 or more times.4,6 STRs are useful in human identity testing due to their variability among members of a population. The forensic community has standardized the use of 13 core STR markers and a number of commercial STR typing kits are now available for examining the DNA variation at these sites (Table 11.1). The 13 core STR markers are CSF1PO, FGA, TH01, TP0X, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11.6 Commercial kits utilize multiplex PCR to simultaneously amplify up to 16 different regions of the human genome with probabilities of a random match between unrelated individuals being smaller than one in a trillion. The commercial STR kits use fluorescently labeled primers to generate PCR products that may be analyzed on multicolor fluorescence detection platforms.7–10 The application of these forensic STRs to criminal DNA databases has greatly expanded their use.6,11

Single nucleotide polymorphisms (SNPs) represent another important class of DNA variation where sequence differences between individuals is examined rather than length variation. A number of different SNP detection and typing technologies exist.6 This chapter will focus on the use of a SNP typing approach that works on the same instrument platform as is commonly used for multicolor fluorescence STR typing, namely the SNaPshot™ assay.

Mitochondrial DNA (mtDNA), which is maternally inherited, can play an important role in many aspects of human identity testing due to the fact that it occurs in higher copy numbers and is more resilient to environmental degradation.12 A desire to gain more information than can be provided by the hypervariable regions of the mtDNA control region has led to a search for informative sites outside the control region around the remaining 15,000 base pairs of the mtDNA genome.13 A set of 10 highly informative sites from around the mtDNA genome has been combined into a multiplex PCR and SNP detection assay that can be detected in high-throughput fashion using multicolor fluorescence and multipapillary instrumentation.

FIGURE 11.1 Sample separation and detection with a multicapillary electrophoresis system. Fluorescently labeled PCR products are separated during electrophoresis and are detected as they pass a region where an excitation laser is shining on the array. Fluorescence is collected onto a charged-coupled device where the various colors can be resolved from one another.
II. MATERIALS AND METHODS

The overall process flows for analyzing both STR and SNP markers with multiplex PCR and multicolor fluorescence detection is illustrated in Figure 11.2. The process involves the use of genomic DNA samples, singleplex or multiplex PCR amplification, sample preparation for multicolor fluorescence detection, data collection on a single capillary or multicapillary array instrument, and data analysis.

A. Genomic DNA Samples

Anonymous human blood samples were purchased from a commercial blood bank (Millennium Biotech, Inc., Ft. Lauderdale, FL) after approval for use through a NIST institutional review board. These samples were extracted using conventional organic extraction methods and quantified via UV spectrophotometry. The extracted genomic DNA samples were diluted to a concentration of 2 ng/µl.

B. PCR Amplification and Typing Assays

1. STR Typing Kits

Manufacturer’s protocols and reagents described in the AmpFiSTR® Blue™ User’s Manual and AmpFiSTR® Identifier™ PCR Amplification Kit User Guide were followed for amplification using

### TABLE 11.1

<table>
<thead>
<tr>
<th>STR Kit Name</th>
<th>Source</th>
<th>STR Loci Included</th>
</tr>
</thead>
<tbody>
<tr>
<td>AmpFiSTR® Blue™</td>
<td>Applied Biosystems</td>
<td>D3S1358, VWA, FGA</td>
</tr>
<tr>
<td>AmpFiSTR® Green 1™</td>
<td>Applied Biosystems</td>
<td>Amelogenin, TH01, TPOX, CSF1PO</td>
</tr>
<tr>
<td>AmpFiSTR® Profiler™</td>
<td>Applied Biosystems</td>
<td>D3S1358, VWA, FGA, Amelogenin, TH01, TPOX, CSF1PO, D5S818, D13S317, D7S820</td>
</tr>
<tr>
<td>AmpFiSTR® Profiler Plus™</td>
<td>Applied Biosystems</td>
<td>D3S1358, VWA, FGA, Amelogenin, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820</td>
</tr>
<tr>
<td>AmpFiSTR® COifler™</td>
<td>Applied Biosystems</td>
<td>D3S1358, D16S539, Amelogenin, TH01, TPOX, CSF1PO, D7S820</td>
</tr>
<tr>
<td>AmpFiSTR® SGM Plus™</td>
<td>Applied Biosystems</td>
<td>D3S1358, VWA, D16S539, D2S1338, Amelogenin, D8S1179, D21S11, D18S51, D19S433, TH01, FGA</td>
</tr>
<tr>
<td>AmpFiSTR® Identifier™</td>
<td>Applied Biosystems</td>
<td>CSF1PO, FGA, TPOX, TH01, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D19S433, D2S1338, amelogenin</td>
</tr>
<tr>
<td>CTTv Multiplex</td>
<td>Promega Corporation</td>
<td>CSF1PO, TPOX, TH01, VWA (vWF)</td>
</tr>
<tr>
<td>FFFL Multiplex</td>
<td>Promega Corporation</td>
<td>F13A1, FES/FPS, F13B, LPL</td>
</tr>
<tr>
<td>GammaSTR® Multiplex</td>
<td>Promega Corporation</td>
<td>D16S539, D13S317, D7S820, D5S818</td>
</tr>
<tr>
<td>PowerPlex® 1.2</td>
<td>Promega Corporation</td>
<td>CSF1PO, TPOX, TH01, VWA, D16S539, D13S317, D7S820, D5S818, amelogenin</td>
</tr>
<tr>
<td>PowerPlex® ES</td>
<td>Promega Corporation</td>
<td>D3S1358, TH01, D21S11, D18S51, vW A, D8S1179, FGA, SE33 (ACTBP2), amelogenin</td>
</tr>
<tr>
<td>PowerPlex® 16</td>
<td>Promega Corporation</td>
<td>CSF1PO, FGA, TPOX, TH01, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Pent D, Penta E, amelogenin</td>
</tr>
</tbody>
</table>

Note: These kits use fluorescent dye-labeled primers and can be run on the ABI 310 or ABI 3100 instrument platforms. The two available 16-plex kits are bolded.
these commercial STR kits, except that reaction volumes were reduced in size from 50 µl or 25 µl to 10 µl. For the results described here, we used typically 10 ng of DNA template and reduced the cycle number from the recommended 28 cycles to 25 cycles because of the higher than recommended amount of DNA template. Thermal cycling was conducted on a GeneAmp 9700 (Applied Biosystems) using the following conditions in 9600-emulation mode (i.e., ramp speeds of 1°C/sec):

95°C for 10 min  
25 cycles: 94°C for 1 min, 59°C for 1 min, 72°C for 1 min  
60°C for 45 min  
25°C hold

2. SNP Typing with the SNaPshot Assay

Primer extension reactions utilizing fluorescently labeled dideoxynucleotides (ddNTPs) were performed using the ABI Prism® SNaPshot™ multiplex system (Applied Biosystems). Prior to performing primer extension reactions, a DNA amplification of 10 regions of the mtDNA genome was performed via multiplex PCR to provide sufficient quantities of the mtDNA amplicons required for probing the polymorphic sites (Vallone et al., manuscript in preparation). The PCR products treated with ExoSAP (USB Corporation, Cleveland, OH) to digest leftover PCR primers and deoxynucleotides (dNTPs) that would interfere with the SNP extension reaction.

SNP extension primers designed to bind to the 5’ region adjacent to each mitochondrial SNP site are listed in Table 11.2. These primers were purchased from Operon Technologies (Alameda, CA). Nonbinding poly-T tails were added to the 5’ end of nine of the extension primers to allow for electrophoretic resolution based on primer length (Table 11.2). Increments of four T nucleotides were used to space the SNP extension products so that they were well resolved from one another.
The tailed extension primers were mixed at concentrations ranging from 0.4 to 10 \( \mu M \) based on empirical optimization of resulting peak heights (Vallone et al., manuscript in preparation).

Each SNP extension reaction contained 1 \( \mu l \) of the mixed tailed SNP extension primers, 4 \( \mu l \) of the multiplex PCR mixture containing the 10 PCR products, and 5 \( \mu l \) of the SNaPshot\textsuperscript{TM} Multiplex Kit reagents (Applied Biosystems, P/N 4323151). Thermal cycling was conducted on a GeneAmp 9700 (Applied Biosystems) using the following conditions in 9600-emulation mode (i.e., ramp speeds of 1°C/sec): 25 cycles: 96°C for 10 sec, 50°C for 5 sec, and 60°C for 30 sec. After thermal cycling 1 unit of SAP (USB Corporation) was added to the reaction and incubated at 37°C for 20 min and then 85°C for 20 min to disable the enzyme activity. This SAP treatment removes unincorporated fluorescent ddNTPs that will migrate in the same region as the labeled primers during electrophoretic separation.

C. Sample Preparation for Data Collection

After multiplex PCR amplification (see Figure 11.2), a 1-\( \mu l \) aliquot of each STR sample was combined with 8.6 \( \mu l \) Hi-Di\textsuperscript{TM} formamide (Applied Biosystems, P/N 4311320) and 0.4 \( \mu l \) GS500 LIZ size standard (Applied Biosystems, P/N 4322682). Typically, the formamide and size standard were combined in the appropriate ratios in sufficient quantities to be dispensed to multiple tubes. After the SNP extension reaction and SAP treatment (see Figure 11.2), SNaPshot samples were prepared in the same manner as the STR samples except that the internal size standard was GS120 LIZ (P/N 4324287).

D. ABI Prism\textsuperscript{®} 310 Genetic Analyzer

Samples amplified with the AmpFISTR Blue kit (Applied Biosystems, P/N 402800) were separated and detected on the ABI Prism\textsuperscript{®} 310 Genetic Analyzer (Figure 11.3). Samples were injected onto the single capillary for 5 sec at 15,000 V. Separations were performed using the standard module GS STR POP4 (1 ml) F at 15,000 V for 24 min with a run temperature of 60°C using POP\textsuperscript{TM}-4 sieving polymer (Applied Biosystems, P/N 402838), 1 X Genetic Analyzer Buffer with EDTA (P/N 402824), and a 47 cm capillary (P/N 402839). The matrix standard set DS-32 (P/N 4312131) was used to generate a spectral matrix for the four dyes 5-FAM\textsuperscript{TM} (blue), JOE (green), NED\textsuperscript{TM} (yellow), and ROX\textsuperscript{TM} (red).
Detection of the multiplex PCR amplified products from the AmpFlSTR® Identifier™ kit (Applied Biosystems, P/N 4322288) was accomplished with the ABI Prism® 3100 Genetic Analyzer 16 capillary array system (Figure 11.4). Manufacturer’s protocols were followed to detect five dyes: 6FAM™ (blue), VIC™ (green), NED™ (yellow), PET™ (red), and LIZ™ (orange). Prior to running the Identifier™ samples, a G5 spectral matrix was established using matrix standard set DS-33 (Applied Biosystems, P/N 4323016). Analysis of samples generated with the SNaPshot™ assay required the establishment of an E5 (5-dye) spectral matrix using the ABI 3100 matrix standard set DS-02 (Applied Biosystems, P/N 4323050). Samples were injected onto the ABI 3100 16-capillary array for 5 sec at 2,000 V or 10 sec at 3,000 V. Separations were performed at 15,000 V for 20 to 45 min, depending on the size of the DNA fragments being analyzed, with a run temperature of 60°C using the 3100 POP™-4 sieving polymer (Applied Biosystems, P/N 4316355) or 3700 POP-5 (P/N 4313087), 1 X Genetic Analyzer Buffer with EDTA (P/N 402824), and a 36 cm array (P/N 4315931).
Following data collection from either the ABI 310 or ABI 3100, samples were analyzed with Genescan® 3.7 (for Windows NT, Applied Biosystems), and allele designations were determined using Genotyper® 3.7 (Applied Biosystems) based on a genotyping macro supplied as part of the commercial STR typing kits. In-house macros were developed for the SNaPshot™ assays. STR allele designations were made based on sizing bin windows of ±0.50 bp and by comparison to allelic ladders. Peak color and size were used to make the SNP call.

III. RESULTS AND DISCUSSION

A. STR TYPING

For each STR marker amplified in a DNA sample, two PCR primers are used to target the region of interest (Figure 11.3, top). One of these primers contains a fluorescent dye at the 5’ end that is used to impart a color to the PCR product when the region between the primers is amplified. Different primer sets can be used to amplify the desired STR repeat region, resulting in somewhat different size ranges for the amplified products.

The PCR product size is converted into the number of repeat units present at the STR region for use in DNA databases through a comparison to an allelic ladder run under the same electrophoretic conditions. Allelic ladders are supplied with commercially available kits and contain commonly observed alleles. All of the alleles in an allelic ladder have typically been sequenced to determine the exact number of repeat units. Therefore, if a PCR product migrates through a gel or capillary at the same rate as one of the alleles in the allelic ladder run under the same conditions (and a high degree of precision has been demonstrated run-to-run), the PCR-amplified allele is
operationally defined as containing the same number of repeats. Genotyping software exists that can compare PCR product sizes between tested samples and allelic ladders and assign a genotype call to each sample for each STR marker tested (Figure 11.3, bottom).

The recent capability for simultaneous detection of five fluorescent dyes (compared to the traditional four dyed chemistries) has expanded the multiplexing space available with each run on a multicolor fluorescence detection instrument and thus increased the potential throughput of these detection systems. A genotyping result with the Applied Biosystems AmpFlSTR® Identifiler™ STR kit run on a 16-capillary array instrument is shown in Figure 11.4. Each set of peaks is labeled with the number of repeats present in the STR markers analyzed. The Identifiler™ kit uses four dyes to label the PCR products blue (6-FAM™), green (VIC™), yellow (NED™), or red (PET™). An internal size standard containing a fifth dye (LIZ™) is used enable sizing of the PCR-amplified alleles in the sample. This kit simultaneously amplifies 15 different STRs from various positions in the human genome and the amelogenin sex-typing marker on the X and Y-chromosomes. All of the 13 core STR markers used within the U.S. for forensic DNA databases are simultaneously amplified in a single multiplex PCR amplification with the Identifiler™ kit. Thus, through a single injection onto a 16-capillary ABI 3100, 256 total data points may be obtained in approximately 45 min.

### B. SNP TYPING

A general schematic of a primer extension reaction using the SNAPSHOT™ multiplex kit is depicted in Figure 11.5. A PCR amplicon, which has been purified to remove unincorporated dNTPs and PCR primers, is the template upon which the primer extension reaction occurs. A SNP extension primer that contains anywhere from 17 to 30 bases complementary to the PCR amplicon binds adjacent to the SNP site of interest (in a 5’ to 3’ direction). Through cycle sequencing, a thermostable polymerase extends the 3’ end of the SNP extension primer by one base unit. Thermal cycling parameters are described in Figure 11.5. Because fluorescently labeled ddNTPs (dideoxy-termina-
tors) are used in the reaction, only a single base is added to the extension primers. The identity of the base added by the polymerase is the complement of the nucleotide at the SNP position. From this base identity one can infer the sequence variant located at the SNP site.

Products of extension reactions are then separated and detected using capillary electrophoresis instrumentation. The instrumentation is capable of detecting each of the four possible dyes (see Figure 11.5) attached to extension primers as well as a fifth dye that is attached to a sizing standard.

When probing more than one marker in a reaction (i.e., using multiple SNP extension primers), additional bases are added to the 5' end of the SNP primer in order to achieve spatial resolution. Typically poly-T tails are used to modify the length of each of the extension primers. This design concept is illustrated in Table 11.2 where the poly-T tails are highlighted from binding region of the SNP extension primer.

The electrophoresis result of a SNaPshot™ primer extension assay designed to probe 10 unique SNP sites located in the coding region of the mitochondrial genome is depicted in Figure 11.6. The data shown in this figure was generated on an ABI 3100 capillary electrophoresis instrument. All 10 fluorescently labeled extension products that range in size from 18 to 58 bases are observed and well resolved. An electrophoretic run of this assay takes approximately 20 min. All four possible sequence variations are represented in this data set. From the “color” of each of the peaks, one can assign the sequence variant present at each of the SNP sites. Traditional sequencing information collected independently (data not shown) confirmed the accuracy of the SNP typing. By combining capillary electrophoresis, multiplex PCR, and extension primer design methods, and the reagents supplied in the SNaPshot™ multiplex kit, it is possible to routinely analyze 5 to 10 SNPs sites.

IV. CONCLUSIONS

The ability to multiplex sample analysis by DNA fragment size, dye label color, and number of capillaries permits high-throughput DNA testing to be performed. The information in Table 11.3 emphasizes that through using higher levels of multiplexing or more capillaries in parallel greater numbers of samples and typing data may be examined more quickly. Even though the overall speed for a single run on a 96-capillary instrument is slower than a single capillary or 16-capillary array system, higher overall throughputs are possible because more capillaries are being run in parallel. As the value of DNA testing continues to grow, the ability to analyze various combinations of STRs or SNPs in parallel and in a high-throughput fashion will be valuable to a number of fields.

www.taq.ir
TABLE 11.3
Comparison of Genotyping Time per Marker Using Multiplex Amplification and Detection with Single and Multicapillary Systems

<table>
<thead>
<tr>
<th>Multiplexed markers</th>
<th>Single Capillary (ABI 310)</th>
<th>16 Capillary Array (ABI 3100)</th>
<th>96 Capillary Array (ABI 3700)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time per run of each capillary</td>
<td>1800 sec (30 min)</td>
<td>2700 sec (45 min)</td>
<td>9960 sec (2 h 46 min)</td>
</tr>
<tr>
<td>Time per sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1800 sec (30 min)</td>
<td>169 sec (2.8 min)</td>
<td>104 sec (1.7 min)</td>
</tr>
<tr>
<td>8</td>
<td>225 sec</td>
<td>21 sec</td>
<td>13 sec</td>
</tr>
<tr>
<td>16</td>
<td>113 sec</td>
<td>10.5 sec</td>
<td>6.5 sec</td>
</tr>
</tbody>
</table>

ACKNOWLEDGMENTS

Funding for this work is gratefully acknowledged from the National Institute of Justice through an interagency agreement with the NIST Office of Law Enforcement Standards. Margaret Kline extracted and quantitated the DNA samples used. Tom Parsons and Mike Coble from the Armed Forces DNA Identification Laboratory helped select the mtDNA sites examined.

REFERENCES

12 The GOOD Assay: A Purification-Free Assay for Genotyping by MALDI Mass Spectrometry

Jörg Tost, Ramón Kucharzak, Doris Lechner, and Ivo Glynne Gut

CONTENTS

I. Introduction................................................................................................................. 121
II. Method of the GOOD Assay........................................................................................ 122
   A. Instrumentation and Suppliers ............................................................................. 122
   B. Software..................................................................................................................... 123
   C. Reagents and Suppliers ............................................................................................. 124
   D. Examples ....................................................................................................................... 125
      1. GOOD Assay with Positive Ion Mode Detection .............................................. 125
      2. The GOOD Assay with Negative Ion Mode Detection....................................... 126
      3. The Simplified GOOD Assay ............................................................................. 126
   E. Interpretation of the Data.............................................................................................. 127
   F. Common Pitfalls and Troubleshooting ...................................................................... 127
      1. SNP Verification ...................................................................................................... 127
      2. Primer Quality.......................................................................................................... 127
      3. DNA Quality ........................................................................................................... 128
      4. Assay Design ........................................................................................................... 128
III. Discussion..................................................................................................................... 128
   A. Budget, Expertise, and Effort Required....................................................................... 128
   B. Other Useful Applications.......................................................................................... 128
   C. Future Developments .................................................................................................... 129

Acknowledgments .................................................................................................................. 129
References ...................................................................................................................................... 129

I. INTRODUCTION

Single nucleotide polymorphisms (SNPs) have gained great interest in the last few years. There are now more than 2 million SNPs available in public databases. Some of the interest in SNPs is founded on the hope that they may help unravel complex genetic traits such as multifactorial disorders through association studies and linkage disequilibrium mapping. However, their power to detect responsible genetic variations is still controversial. SNPs have several advantages compared to other markers like short tandem repeats (STRs). They are highly abundant, are binary in
nature, can be anchored in intragenic regions, and are relatively easy to analyze. A wide variety of methods using diverse detection devices has been presented for all scales of genotyping studies, each having its advantages and pitfalls (see References 5 and 6, and Chapter 10 by Kalim U. Mir and Jiannis Ragoussis). MALDI mass spectrometry as a detection device has been successfully applied for SNP genotyping using allele-specific hybridization, ligation of allele-specific oligonucleotides, allele-specific cleavage of oligonucleotides, and primer extension to generate allele-specific products. In principle, MALDI mass spectrometry is ideally suited for high-throughput genotyping studies due to the rapid and accurate nature of its data accumulation. Mass, a physical property of the product rather than a signal deriving from a tag of the product, is read out. The multichannel detecting capabilities of mass spectrometers allow multiplexing.

In the beginning, analyzing nucleic acids with mass spectrometers proved to be more difficult than anticipated. Signal intensity is about 100 times lower than for peptides of the same mass. Cations, especially of sodium and potassium, form adducts with DNA resulting in diffuse signals and can only be removed by stringent purification protocols. However, purification contributes largely to the genotyping costs and makes automation more cumbersome. The GOOD assay circumvents purification with the help of a chemical modification strategy. Negative charges from the backbone are neutralized and a single positive or negative charge is introduced chemically, rendering the oligonucleotide insensitive to the formation of adducts and thereby increasing the detection sensitivity about 100-fold, equaling that of peptides. This strategy was implemented in the GOOD assay (Figure 12.1). The regular GOOD assay consists of five steps: (1) a stretch of DNA including the SNPs of interest is amplified by PCR, (2) remaining dNTPs are removed by a Shrimp Alkaline Phosphatase (SAP) digest, (3) the SNPs are queried using a primer extension reaction, (4) extension primers are reduced to the core-sequence by phosphodiesterase II digestion, and (5) charges from the backbone of the remaining extension primer are alkylated using iodomethane. All reaction steps are simple additions of reagents into the same tube or wells of a microtitre plate followed by incubations. Finally, samples are transferred onto the MALDI target without any purification whatsoever.

II. METHOD OF THE GOOD ASSAY

The GOOD assay has been applied in several variations. Each variant has its specific application and advantages, and which one is best suited depends on the genotyping experiment. The GOOD assay with positive ion mode detection utilizes modified primers containing an amino-functionalized and charge-tagged base bracketed by two phosphorothioate bridges for the primer extension reaction (Figure 12.2A). Because charge tags can be used to shift allele masses, the assay is well suited for multiplexing.

The regular GOOD assay with negative ion mode detection avoids the expensive amino-modified oligonucleotides for primer extension (Figure 12.2B). In return, the possibilities for multiplexing are limited and detection of the products in the negative ion mode of the mass spectrometer is slightly more difficult.

The simplified GOOD assay is the most recent development, reducing the five-step procedure of the regular GOOD assay to three liquid handling steps, thus saving time and effort. The SNPs are queried using extension primers containing methylphosphonates extended with ddNTPs that also contribute the negative charge for sensitive detection (Figure 12.2C). Excellent signal-to-noise is achieved with the simplified GOOD assay.

A. INSTRUMENTATION AND SUPPLIERS

Mass spectrometric analyses were done in linear mode on either a Reflex® or an Autoflex® MALDI mass spectrometer (Bruker Daltonik, Bremen, Germany) equipped with a Twister robot (Zymark, Hopkinton, MA). Typical acceleration potentials were ±18 kV and ion extraction was delayed by 200 nsec. All liquid handling steps for the GOOD assay are done on a
**The GOOD Assay**

**Regular**

1. PCR
2. SAP
3. Primer extension
4. PDE digest
5. Alkylation
6. Thermocycling
7. Incubation

**Simplified**

1. PCR
2. Primer extension
3. PDE digest
4. Incubation

**MALDI ANALYSIS**

**FIGURE 12.1** The flowchart shows the procedure for the different variants of the GOOD assay. The regular GOOD assay can be performed with either positive or negative ion mode detection. The simplified GOOD assay in negative ion mode reduces the five-step procedure to three steps.

BasePlate® Robot from The Automation Partnership (Royston, U.K.), with special programming for the protocols. This robot is equipped with a 96-tip liquid handling head. PCR protocols are developed in 3 µl on a 384 Gradient Mastercycler (Eppendorf, Hamburg, Germany). All thermocycling procedures are then carried out in Primus Thermocyclers from MWG Biotech (Ebersberg, Germany).

**B. SOFTWARE**

PCR primers are designed using the PRIMER 3 software (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). Oligonucleotides for primer extension are determined with software developed in-house called SNIP machine (Pierre Lindenbaum, http://www.cng.fr). This software chooses the best position and direction for the primer, minimizing secondary structures in the extension primer and the risk of interactions with other primers when multiplexed. Chapter 13 by Maida Remm, Ants Kurg, and Andres Metspalu contains detailed discussion of software for the design of primers in multiplex PCR and single nucleotide primer extension assays.

MALDI spectra are recorded automatically. Spectra are accepted if the intensity of the signal at a given position fulfills the set requirements of signal-to-noise ratio and resolution. Analysis of the raw data is performed using the Genotools® SNP manager software.13
Primers for PCR and primer extension (containing one phosphorothioate bridge) for negative ion mode detection were purchased from MWG (Ebersberg, Germany). Primers for primer extension containing the amino-Modified phosphoramidites and two phosphorothioate bridges (for the GOOD assay with positive ion detection) were synthesized by Biotez (Berlin, Germany). Extension primers for the simplified GOOD assay containing methylphosphonate bridges were from Eurogentec (Liège, Belgium). Shrimp Alkaline Phosphatase (SAP) and ThermoSequenase were purchased from Amersham (Little Clafont, U.K.), TMA31FS from Roche Molecular Systems (Alameda, CA), Phosphodiesterase II (from calf spleen) from Sigma (St. Louis, MO), and α-S-ddNTPs from Biolog (Bremen, Germany). All reagents are stored at –20°C. Mastermixes (without enzymes and nucleotides) can be prepared ahead and stored at –20°C. The MALDI matrix α-cyano-4-hydroxycinnamic acid methyl ester, charge tag reagents, and the β-cyanoethyl phosphoramidites are available from Bruker Saxonia Analytik GmbH (Leipzig, Germany).

Charge tagging of the amino-Modified oligonucleotides is performed using different trialkylammoniumalkyl-α-hydroxy succinimidyl esters, prepared according to the procedure described by Bartlet–Jones et al. The trimethylammonium-butyl-, -pentyl-, and -hexyl...
esters are available commercially. In addition, we have synthesized the ethyldimethylammonium-, the diethylmethylammonium-, and the triethylammonium-hexyryl esters. Of the total amino-modified oligonucleotide, 15 nmol are dissolved in 30 μl 1% TE, mixed with 1.5 μl triethylammoniumhydroxycarbonat buffer (pH~8.0) and 24 μl of a freshly prepared 0.75 to 1.5% solution of the charge tag reagent in water, incubated at 0°C for 30 min. The reaction mixture is lyophilized, resuspended in 15 μl 300 mM diammonium acetate, and the oligonucleotide purified by ethanol precipitation. The precipitate is resuspended in 30 μl of ddH₂O, the concentration measured by UV spectroscopy at 260 nm and adjusted to a final concentration of 25 pmol/μl.

D. EXAMPLES

1. GOOD Assay with Positive Ion Mode Detection

This assay is currently used in our high-throughput genotyping facility. Except for the PCR and the amount of extension primer used, conditions are the same for all experiments. As an example, we show a duplex genotyping experiment for two SNPs (G-486T and C-410T) in the Caveolin 3 gene. Figure 12.3 shows spectra for three different DNA samples.

**PCR:** For amplification of the DNA fragment (274 bp) encompassing the two SNPs of the Caveolin 3 gene, 2.25 pmol each of forward (5’-AGG TCA CCA AGC AGG) and reverse (5’-CTG GAA TTT GCC CCA CTG) primer and 1 μl of DNA (5 ng/μl) as template were used. The reaction conditions were 60 mM Tris-SO₄ (pH 8.9), 18 mM (NH₄)₂SO₄, 1.0 mM MgSO₄, 100 μM dNTPs, and 0.2 U Platinum® Taq DNA Polymerase High Fidelity (Invitrogen, Life Technologies [Carlsbad, CA]) in a final volume of 3 μl. The reaction mixture is covered with 3 μl mineral oil to avoid evaporation. Reactions were denatured for 4 min at 94°C and then thermocycled for 30 sec at 94°C, for 20 sec at 56°C, and for 30 sec at 72°C, repeating the cycle 35 times, followed by a final extension of 4 min at 72°C.

**SAP digest:** 2 μl containing 0.25 U of shrimp alkaline phosphatase (SAP) in 44 mM Tris buffer (pH~8.0) were added to the PCR reaction and incubated for 1 h at 37°C, followed by inactivation of the enzyme at 90°C for 10 min.

![Figure 12.3](http://www.taq.ir)
Primer extension reaction: 5 pmol of the primer querying position G-486T (5’ ACA GAT GGG GAC AGA G₃p(T₆)₃ G, pt = phosphorothioate bridge, CT = charge tag), 5 pmol for position C-410T (5’ -TAA CCC CAC TTC CCT C₃p(C₃)₃ A), and 1 U ThermoSequenase were added. The Mg²⁺ concentration was adjusted to a final concentration 4.7 mM, α-S-ddCTP and α-S-ddGTP to 28.5 μM and α-S-ddTTP to 57 μM. The final reaction volume was 7 μl. The reaction mixture was thermocycled with an initial denaturing step of 2 min at 94°C followed by 30 cycles of 15 sec at 94°C, 20 sec at 56°C, and 20 sec at 72°C.

PDE digestion: 2 μl containing 0.5 μl of 0.5 M acetic acid and 6.5 × 10⁻³ U of phosphodiesterase II in 0.2 M diammoniumhydrogencarbonate buffer were added to the reaction and incubated for 60 min at 37°C.

Alkylation reaction: A mixture of 11 μl acetonitrile, 0.5 μl 2 M triethylammonium-hydrogencarbonate buffer (pH≈7.5), 2.5 μl of 2 M Tris-HCl buffer (pH≈8), and 5.25 μl iodomethane was added and incubated at 40°C for 25 min. Upon cooling, a biphasic system was obtained. 10 μl of water were added and allowed to stand for 5 min at room temperature. 5 μl of the upper layer were taken off and diluted with 10 μl of 40 % acetonitrile.

Sample preparation: 0.5 μl of a 1.5 % solution of α-cyano-4-hydroxy-cinnamic acid methyl ester in acetone was used to prepare the matrix and 0.5 μl of sample was spotted on top with the BasePlate robot.

Mass spectrometric analysis: Spectra were recorded automatically on a Bruker AutoFlex® MALDI time-of-flight mass spectrometer using the AutoXecute® software. Spectra were recorded in positive ion linear time-of-flight mode.

2. The GOOD Assay with Negative Ion Mode Detection

In the negative ion detection mode, protononation of the product by the matrix would weaken the signal. Therefore, salt concentrations have to be reduced compared to assays with positive ion detection. As an example, we present the protocol for the assay querying SNP WI1126 (GenBank 719980).

PCR: 2.25 pmol of forward (5’-CCA GAT TGT TTT CCC AGC) and reverse primer (5’-GAT GCC AAA CTA ATA GTG C), and 1 μl of DNA (5 ng/μl) were used to amplify a product of 173 bp. 0.2 U of Platinum® Taq DNA Polymerase (Invitrogen, Life Technologies [Carlsbad, CA]) were used in 3 μl volume containing 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1 mM MgCl₂, and 100 μM dNTPs. The reaction mixture was denatured for 4 min at 94°C followed by 30 cycles of 30 sec at 94°C, 45 sec at 56°C, and 30 sec at 72°C subsequently followed by 4 min at 72°C.

SAP digest: 2 μl containing 0.25 U of SAP in ddH₂O were added to the PCR reaction and incubated for 1 h at 37°C followed by denaturation at 90°C for 10 min.

Primer extension reaction: 15 pmol of the primer querying position WI1126 (5’-CAA GAA ATA AAA GAC AGG T₃p(CC)) and 1 U ThermoSequenase were added. The Mg²⁺ concentration was adjusted to a final concentration of 3.3 mM and 0.57 μM α-S-ddATP and 0.57 μM α-S-ddGTP were added. The reaction volume was 7 μl. After an initial denaturation step of 2 min at 95°C, the reaction mixture was thermocycled using 30 cycles of 10 sec at 95°C, 15 sec at 56°C, and 10 sec at 72°C.

PDE digestion, alkylation reaction, and sample preparation: These steps were performed as described above except that the Tris buffer in the alkylation was replaced by 2 M triethylammonium-hydrogencarbonate buffer (2.75 μl final volume).

Mass spectrometric analysis: Spectra were recorded automatically on a Bruker Reflex® time-of-flight mass spectrometer in negative ion linear mode with the AutoXecute® software.

3. The Simplified GOOD Assay

The SNP G+58A in the Platelet/endothelial cell adhesion molecule (PECAM) gene is presented as an example.12

www.taq.ir
**PCR:** 2.25 pmol of forward (5′-CAT TTT GCA TTT CTC TCC ACC) and reverse primer (5′-GCA GGG CAG GTT CAT AAA TAA G) and 1 μl of DNA (5 ng/μl) were used to amplify a product of 218 bp. 0.2 U of Platinum® Taq DNA Polymerase High Fidelity were required in 3 μl volume containing 60 mM Tris-SO₄ (pH 8.9), 18 mM (NH₄)₂SO₄, 1 mM MgSO₄, and 100 μM dNTPs. The reaction mixture was denatured for 4 min at 94°C followed by 30 cycles of 30 sec at 94°C, 45 sec at 63°C, and 30 sec at 72°C followed by 4 min at 72°C.

**Primer extension reaction:** 0.8 U TMA31FS polymerase and 4 pmol of the extension primer for the SNP G+58A (5′-ATG TTC CGA GAA GAA CAG mpAmpT, mp = methylphosphonate bridge) were added to the 3 μl PCR. Because methylphosphonate bridges are already charge neutral, no charge neutralization is required. The Mg²⁺ concentration was adjusted to a final concentration 6.6 mM, ddATP and ddGTP to 30 μM in a final reaction volume of 5 μl. After an initial denaturing step of 1 min at 94°C, the reaction mixture was thermocycled (30 cycles) with 94°C for 15 sec, 54°C for 20 sec, and 72°C for 20 sec. The heating rate between annealing and extension step was 0.5°C/sec.

**PDE digestion:** 2 μl containing 0.5 μl of 0.5 M acetic acid and 6.5 ¥ 10⁻³ U of phosphodiesterase II in 0.2 M diammoniumcitrate buffer were added to the reaction and incubated for 60 min at 37°C.

**Sample preparation and mass spectrometric analysis:** These steps were performed as described above for the GOOD assay in negative ion mode.

**E. INTERPRETATION OF THE DATA**

The raw data is analyzed automatically using the Genotools® software. Its algorithm for allele-calling relies on the determination of intensity and deviation from the calculated mass (tolerance) of the signals at given points. Spectra with low reliability are batched for visual inspection. The analyzed data is then checked for deviations from the Hardy–Weinberg equilibrium. Hardy–Weinberg equilibrium calculates the expected genotype frequencies under the assumption of random assortment and absence of population admixture in the tested DNAs. Plates deviating significantly from the equilibrium are inspected visually. If several SNPs in a gene are characterized, haplotypes (i.e., the phase of several SNPs on a parental allele) are mathematically deduced. The raw data of infrequent haplotypes (in the examined population) is again visually inspected. Verified data is then correlated with phenotypes in question.

**F. COMMON PITFALLS AND TROUBLESHOOTING**

1. **SNP Verification**

There are about 4 million SNPs in the public databases to date, but only a small part of them has been verified. Further information about allele frequencies in different ethnic backgrounds and linkage disequilibrium patterns is often missing. Many of the apparent polymorphisms in the public domain have been deduced in silico and may have been caused by sequencing errors or duplications wrongly assigned to the same locus. Therefore, SNPs have to be confirmed by sequencing or similar techniques in a subset of the population which will be characterized. We routinely sequence between 32 and 96 individuals for confirmation. In this verification step we also often find hitherto unknown SNPs. This underlines the importance of verification as these SNPs may cause bias in the amplification or primer extension reactions if the primers used are fully complementary only to one allele.

2. **Primer Quality**

To a great extent, the quality of amplification and extension primers determines the success rate of an assay. We routinely perform quality control on all primer stocks using the MALDI mass spectrometer. All primers are stored frozen at −20°C. Dilutions are prepared fresh from stock. In general, stocks are frozen and thawed less than 10 times before they are exhausted. So far, we have not observed degradation of primers over time.
3. DNA Quality

It is worth making sure that template DNA is of good quality and at a precise concentration. Variation of quality and quantity has great impact on the quality of results. In our experience, phenol/chloroform extracted DNA gives the best results. DNA integrity over time and, thus, reproducibility are good. DNA stocks are stored at –20°C. Daughter plates are prepared with 200 µl of DNA per well at a concentration of 5 ng/µl. These are used to prepare granddaughter (production) plates with 1 µl of DNA per well. Daughter plates are used up within 10 times freezing and thawing. Granddaughter plates are stored at –20°C. They can be stored over long periods of time without decrease of genotyping quality.

4. Assay Design

The GOOD assay is robust, even accepting DNA template of low quality if optimized accordingly. Careful design of assays helps to minimize problems. MALDI mass spectrometry is sensitive to impurities and, as no purification step is included, the reagents of each step have to be compatible with the analysis. There are additives in the storage and reaction buffers of several DNA polymerases that strongly interfere with the analysis and hence should be avoided. In a multiplex experiment, masses of the extension products should differ by at least 5 Da to allow clear assignment of peaks. On rare occasions, a peak corresponding to an incomplete phosphodiesterase digestion of the nonextended primer might be observed. This is sequence dependent; if there are three cytosines at the 3’ end of the extension primer, PDE digestion can be incomplete. This can be avoided by querying the other strand of the PCR product or by introducing a third phosphorothioate bridge into the extension primer. In principle, assays should always be designed so that a potential peak resulting from an incomplete digestion can clearly be distinguished from the signals corresponding to the alleles of the SNP.

III. DISCUSSION

A. BUDGET, EXPERTISE, AND EFFORT REQUIRED

Setting up SNP genotyping by MALDI mass spectrometer requires a significant investment of capital due to the prices of mass spectrometers and robotics. On the other hand, the reagent cost per SNP measured is lower than those for competing methods using fluorescent detection. In large-scale studies, the reagents contribute the major part of the cost, whereas at medium- to low-throughput instrumentation and labor are the major contributors. Miniaturization and higher multiplexes further decrease the costs. We are currently running SNP genotyping production with two people, producing about 20,000 genotypes per day. Organization is one of the biggest problems when running SNP genotyping at high-throughput, an aspect discussed in more detail in Chapter 10 by Kalim U. Mir and Jiannis Ragoussis.

B. OTHER USEFUL APPLICATIONS

Our assay has been successfully applied for SNP genotype fingerprinting for tracing farm animals (cattle, sheep, pigs). PCR for this application uses crude DNA extract obtained by proteinase K digestion of tissue samples taken during ear tagging. This makes expensive DNA extraction unnecessary.

Short-range sequencing can be done along the same line as the GOOD assay by integrating a set of α-S-dNTPs and α-S-ddNTPs. We have applied short-range sequencing to genotyping of a length polymorphism in the Interleukin 6 promoter (Kucharzak et al., unpublished). The polymorphism contains a polyA/polyT sequence 3’-AATCTGGTC ACTGA (n = 8–10) T(n = 10–12)-CAAAAAACATAGC-5’, with the following known combinations of polyA/polyT: A8-T12, A9-T11, A10-T10.
The GOOD Assay: A Purification-Free Assay for Genotyping by MALDI Mass Spectrometry

To genotype this polymorphism, the primer extension step of the GOOD assay is carried out with sense and antisense primer extension integrating $\alpha$-S-dATP and $\alpha$-S-ddTTP for the extension and termination. The extension primer is anchored outside the polyA–polyT sequence and reaches over the first eight As. The extension is terminated on the first T. The compound information from the sense and antisense extension primer is used to assign genotypes. Genotyping of this polymorphism is impossible by any other genotyping method as the two alleles are visualized at different masses. The only way this polymorphism can be analyzed is by a cumbersome PCR, cloning, and sequencing strategy.

A way to decrease costs is by genotyping pools of DNA. We have used the GOOD assay for quantitative SNP genotyping. The limit of detection for the minor allele is around 5% and allele frequencies can be determined with an accuracy of 3%. However, by this strategy rare alleles can be missed and further information, like individual genotypes, is lost.

C. FUTURE DEVELOPMENTS

We are currently extending our studies to plants, in which SNPs analysis is complicated by gene duplications and different ploidy. In plants, examples of triallelic SNPs are found. Using the MALDI mass spectrometer for analysis, the three alleles are easy to distinguish (Giancola et al., unpublished).

Haplotypes increase the content of information obtained from genotyping experiments and may be more informative for pharmacogenetic assessment. Haploid DNA stretches can be isolated by a highly allele-specific PCR, and the phase of SNPs is then determined in a multiplex experiment, thereby making use of the multichannel capabilities of mass spectrometers.

ACKNOWLEDGMENTS

We are grateful for support from the French Ministry of Research and the European Commission.

REFERENCES


www.taq.ir
13 Primer Design for Large-Scale Multiplex PCR and Arrayed Primer Extension (APEX)

Maido Remm, Ants Kurg, and Andres Metspalu

CONTENTS

I. Introduction ................................................................. 131
II. Bioinformatics Support for Large-Scale Genomic PCR ........................................... 132
   A. Optimal Parameters for PCR Primer Design .................................................... 132
   B. Genome Test ...................................................................................................... 132
   C. Multiplexing PCR Primers .............................................................................. 134
III. PCR Protocol ................................................................. 135
IV. SNP Genotyping Using APEX .................................................. 135
V. Discussion of Problems and Practical Hints ....................................................... 138
   A. Practical Hints for the Primer Design ............................................................ 138
   B. Practical Hints for the Genome Test .............................................................. 138
   C. Practical Hints for APEX Genotyping ........................................................... 139
VI. Conclusions and Future Directions ................................................................. 139

References .................................................................................................................. 139

I. INTRODUCTION

Variation in the human genome plays a central role in susceptibility to disease and the results of drug therapy; understanding it will be a cornerstone of future “personalized medicine.” However, detailed structure of the human genome is still not fully understood. Recently, four studies have been published demonstrating the fine structure of the human chromosome 21, 22, and selected autosomal regions across the human genome. Linkage disequilibrium (LD) was found to be highly variable across the chromosomes studied, and there is no reason to think that the rest of the genome is very much different. More surprising was the finding that LD blocks have been very stable over the course of the human history, and only 3 to 5 common haplotypes per block may account for about 80% of the Caucasian population. Therefore, only 2 to 3 well selected SNPs (“haplotype tag SNPs” or htSNPs) are needed per block to reveal the LD block structure of the human genome. Although this still means that 300,000–400,000 SNPs have to be genotyped to get the full LD structure, we estimate that only 35,000 htSNPs are needed to represent LD structure of 90% of the exons in the genome. This is already within the reach for the today’s genotyping technology. Here we describe PCR and genotyping protocols that allow us to solve problems at genomic scale with thousands of SNPs.
II. BIOINFORMATICS SUPPORT FOR LARGE-SCALE GENOMIC PCR

Mass-genotyping platforms now allow testing of up to 100,000 SNPs from each individual. However, most of these methods require PCR amplification to achieve sufficiently strong signals. Amplification of thousands of fragments from the human genome is not a trivial task. Automatic methods for primer design are necessary to do it within a reasonable time limit. Achieving high primer design quality with automatic methods is typically harder than with manual design. However, automatic methods are able to surpass the speed and quality of manual design if the primer parameters are properly selected. We have collected and analyzed PCR data from the genotyping experiment where 1279 different markers were amplified and genotyped from the chromosome 22. The correlation between PCR quality and different primer properties was estimated.

A. OPTIMAL PARAMETERS FOR PCR PRIMER DESIGN

Approximately 300 primers from our set were designed manually; the rest were designed automatically with the program PRIMER3.7 Candidates that matched common repeat sequences, such as Alu, Line, and MIR, were eliminated from the set. Of the remaining oligonucleotides, 714 primer pairs that were experimentally verified more than 10 times were selected for correlation analysis. PCR results with missing band, multiple bands, band of wrong size, or smeared band were counted as negative result. We then analyzed the effect of various primer design parameters on the fraction of positive results (Table 13.1). Our results are similar to a statistical analysis reported by David Cox and colleagues.8 In general, longer and AT-rich primers work slightly better in genomic PCR. A “GC-clamp” is often designed into the 3’ end of primers to achieve higher specificity. We did not observe any benefits; to the contrary, primer pairs where both primers had a GC-clamp showed 10% decrease in average PCR quality, possibly because of enhanced primer-dimer formation or unspecific binding in the genome.

B. GENOME TEST

In addition to traditional parameters, we also tested the relationship between PCR quality and the number of primer binding sites and the number of products in the human genome sequence. A string consisting of the last 15 nucleotides from the 3’ end was used to search for matches in the

<table>
<thead>
<tr>
<th>Primer Property</th>
<th>Effect on Assay Quality</th>
<th>Suggested Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>Weak</td>
<td>Both primers &gt; 23 nt</td>
</tr>
<tr>
<td>GC%</td>
<td>Weak</td>
<td>Average GC 35–55%</td>
</tr>
<tr>
<td>GC% of the 3'-half</td>
<td>Medium</td>
<td>Average GC 30–65%</td>
</tr>
<tr>
<td>GC% within last 3'-nucleotides of both primers</td>
<td>Weak</td>
<td>Both primers ending with A or T</td>
</tr>
<tr>
<td>GC% within last two 3'-nucleotides of both primers</td>
<td>Weak</td>
<td>Average GC Σ 50%</td>
</tr>
<tr>
<td>GC% within last three 3'-nucleotides of both primers</td>
<td>Weak</td>
<td>Average GC% Σ 50%</td>
</tr>
<tr>
<td>GC% within last four 3'-nucleotides of the primer</td>
<td>Medium</td>
<td>Average GC Σ 50%</td>
</tr>
<tr>
<td>GC% in last half</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>GC% of the whole primer</td>
<td>Medium</td>
<td>100–400 nt</td>
</tr>
<tr>
<td>PCR product size</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
human genome. Only sites with 100% identity to primer sequence were counted. PCR product was counted if sense–antisense primer binding sites were found in the genome within 1000 bp from each other. The results in Table 13.2 show convincingly that testing of the primers against the genome is essential for high quality genomic PCR.

We call the process of determining the genomic locations of PCR primer binding sites the genome test. If all possible binding sites for both primers are known, counting the number of alternative PCR products is trivial. Although the results in Table 13.2 show that primers that could generate up to five products from the genome might work in a PCR reaction, it is wise to avoid primer pairs that generate more than one product. For example, in genotyping experiments, any additional PCR product could interfere with the genotyping signal from the real SNP marker, so that the observed signal is, in fact, a mixture from two different PCR products. Although this situation might be detectable as heterozygote excess in the Hardy–Weinberg test, the results from such marker cannot be used in further analyses.

Similarity searches with the sequence to be amplified, in order to avoid pseudogenes or repeated motifs, have been in practice for a long time. However, such similarity search typically returns a local alignment only with the best matching area. From this local alignment it is hard to make sure that the chosen primers really are unique and generate a unique product from the genome. Furthermore, running BLAST on NCBI server against the typical “nr” database is slow, can be difficult to interpret and does not allow proper estimation of the number of products. Our solution to the problem is to directly localize and enumerate all possible binding sites in the genome. We have tested different string and similarity search programs in order to find the most efficient solution.

A search against the 3 × 10^9 nucleotides of the human genome cannot be performed on a simple desktop computer. Even if only matches with 100% identity (no gaps or mismatches) are allowed, it requires a very large memory size. There is always a certain trade-off between program speed and memory usage, such that slower programs may cope with 260 MB of RAM, but faster programs demand 1 GB or more RAM. We tested widely available computer programs BLAST, MEGABLAST, and SSAHA, among others. Although SSAHA, unlike BLAST and MEGABLAST, is not able to find matches that contain gaps and mismatches, the overall result (number of products found from the genome) is almost always identical. For computers with at least 1 GB RAM we recommend using MEGABLAST or SSAHA. Other programs (including BLAST) are unacceptably slow when analyzing more than 50 primer pairs. The speed of BLAST can be increased by running it over the Internet, for example on the NCBI server. However, the databases available at NCBI site do not allow efficient determination of the genomic locations of primers. We have used assembled genomic data from the ENSEMBL database, either with all contigs from the golden path, or for individual chromosomes separately.

One important parameter is the length of the string (“word length”) from the primer’s 3' end chosen for searching binding sites. Ideally, this should be the length of primer that is sufficient for

<table>
<thead>
<tr>
<th>Primer Property</th>
<th>Effect on Assay Quality</th>
<th>Suggested Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of binding sites in the genome (higher value of both primers)</td>
<td>medium</td>
<td>&lt; 100 binding sites</td>
</tr>
<tr>
<td>Number of binding sites in the genome (lower value of both primers)</td>
<td>strong</td>
<td>&lt; 10 binding sites</td>
</tr>
<tr>
<td>Number of PCR products generated from the genome</td>
<td>strong</td>
<td>&lt; 5 products</td>
</tr>
</tbody>
</table>

Note: Our current data suggest that two primers with an intermediate specificity (i.e., 30 and 20 binding sites) are worse than a pair where one has a very low, one a very high specificity (i.e., 200 and 5 binding sites). A more thorough analysis will be published in the future.
specific binding at a given $T_m$ and given salt, primer, and template concentrations. The sufficient word length might be different for different primers — longer for AT-rich primers and shorter for GC-rich primers. In practice, it is easier to use equal word length for all primers. We have tested different word lengths on our test set mentioned above to see which gives the best separation between “good” and “bad” primers. Our results show that word lengths between 15 and 18 nucleotides give the best predictive power in the genome test. The results shown in Table 13.2 are calculated with word length 15.

We have implemented these findings in the computer program GenomeTester. It is able to determine genomic locations of up to 100,000 primer pairs within few minutes. Single primer pair can be analyzed in a few seconds (Table 13.3). A demo version of the program is available for testing over the WWW.

### C. Multiplexing PCR Primers

Amplification of many marker regions simultaneously in a multiplex reaction offers significant savings in template DNA, time, and money. Setting up a multiplex PCR with consistent quality is not trivial. Often, significant efforts are made to optimize reagent sources and concentrations in multiplex PCR. However, testing for primer–primer and primer–product interactions in each group is equally important. Primer dimer formation can be avoided by designing identical nonpalindromic dinucleotides to the end of each primer, but this is often difficult to achieve for large number of primers. We have created a computer program, MULTIPLX, which tests all primer pairs for interactions, including dimer formation, and automatically generates compatible primer groups for multiplex PCR. In principle, the primers in one multiplex group should conform to the same criteria as a primer pair in a single PCR reaction — they should not form primer dimers with each other and they should not bind to each other’s products. Therefore, the same criteria that are used to avoid primer–primer interactions in PRIMER3 could be appropriate for checking primer–primer interactions in MULTIPLX. The program PRIMER3 uses two parameters to check for possible primer–primer interactions: PRIMER_SELF_ANY and PRIMER_SELF_END. These are calculated as local and global alignment scores between two primers or between primer and its complement. We added two more parameters: PRIMER_PRODUCT_ANY and PRIMER_SELF_END1. PRIMER_PRODUCT_ANY is similar to PRIMER_SELF_ANY except that it calculates the highest score of the primer–product local alignment, instead of primer–primer local alignment. With rather stringent set of parameters we have been able to achieve average 8-plex and maximum 16-plex PCR without any decrease in the genotyping call rate. Our recent tests indicate that these parameters can be further relaxed (PRIMER_SELF_ANY = 8; PRIMER_SELF_END = 6; PRIMER_PRODUCT_ANY = 12) achieving 50-plex PCR without significant loss in quality.

---

**TABLE 13.3**

<table>
<thead>
<tr>
<th>Program</th>
<th>Time to Test 100 Primer Pairs</th>
<th>Time to Test 1000 Primer Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLAST2</td>
<td>7000 sec</td>
<td>—</td>
</tr>
<tr>
<td>MEGABLAST</td>
<td>700 sec</td>
<td>7105 sec</td>
</tr>
<tr>
<td>SSAHA</td>
<td>1700 sec</td>
<td>2540 sec</td>
</tr>
</tbody>
</table>

*Note:* 100 or 1000 PCR primer pairs were tested against the human genome to find all possible binding sites and all possible products. The tests were run on Pentium III, running Linux, with 2 GB of RAM and SCSI hard drives in RAID5 configuration.
III. PCR PROTOCOL

DNA samples from 5 CEPH families and 51 random individuals from the Estonian population were screened for 1279 SNP loci on chromosome 22. The corresponding PCR products were amplified in 384-well plates using two different touchdown PCR programs. In the dNTP mixture 20% of dTTP was substituted by dUTP.

For GC-rich products (GC% > 50%), covering 1045 SNP loci, the program consisted of:

- 5 min at 95°C predenaturation,
- 8 cycles of 20 sec at 95°C, 25 sec at 68–60°C (at 1°C per cycle), 30 sec at 72°C
- 12 cycles of 20 sec at 95°C, 25 sec at 60°C, 30 sec at 72°C
- 4 cycles of 20 sec at 95°C, 25 sec 60–56°C (at 1°C per cycle), 30 sec at 72°C
- 5 min at 72°C final extension.

The second program for the remaining 234 SNP loci (GC% < 50%) was:

- 5 min at 95°C predenaturation
- 15 cycles of 20 sec at 95°C, 25 sec at 58–50°C (at 0.5°C per cycle), 30 sec at 72°C (30 sec)
- 15 cycles of 20 sec at 95°C, 25 sec at 50°C, 30 sec at 72°C
- 5 min at 72°C final extension

The amplification products were concentrated and purified by ethanol precipitation. DNA fragmentation and functional inactivation of unincorporated dNTP was achieved by addition of thermolabile uracil-N-glycosylase (Epicentre Technologies, Madison, WI) and shrimp alkaline phosphatase (SAP) (Amersham Biosciences, Piscataway, NJ).

IV. SNP GENOTYPING USING APEX

During recent years, single nucleotide primer extension methodologies have been widely accepted as the method of choice for high-throughput genotyping of SNPs using different assay formats and detection platforms\textsuperscript{17–23} (see Chapter 12 by Jöng Tost on the GOOD assay and Chapter 17 by Osman El-Maarri on quantitative DNA-methylation analysis). In our lab, we have developed the Arrayed Primer EXtension (APEX) method and adjusted it to large-scale genotyping (Figure 13.1). The method is based on hybridization of fragmented template DNA with short 25-mer oligonucleotides, which have been arrayed on the solid surface. Alternatively, generic tags have been used to identify the products of primer extension reactions either in microarray format carrying complementary oligonucleotide tags\textsuperscript{24} or arrays of microspheres carrying unique complementary “tag” sequence for capture of primer extension products.\textsuperscript{25} Hybridization is followed by single base extension of the immobilized oligonucleotide primers with four fluorescently labeled dye terminators. All primers used in the APEX assays are based on the wild-type DNA with their 3' ends complementary to bases just before the called position.
APEX is utilising two oligonucleotides, one for sense and other for antisense direction per one analyzed basepair combining both high information content of oligonucleotide array and specificity of molecular recognition by DNA polymerase. All four dye-terminators are used in the same reaction, allowing simultaneous evaluation of all possible nucleotide changes (Figure 13.2). APEX is capable for identification of different types of mutations and polymorphisms and can also be used for gene resequencing.27

We used 24 × 60 mm glass microscope slides silanized with aminosilane and coated with phenylenedisothiocyanate as the support material for microarrays (Asper Biotech Ltd., www.asperbio.com). The 25-mer oligonucleotides, equipped with C12 amino linker at their 5' ends, were obtained from MWG-BIOTECH AG (Ebersberg, Germany). Primers were diluted to 50 μM final concentration in 100 mM sodium carbonate/bicarbonate buffer (pH 9.0) and spotted onto activated glass surface with Virtek-Chip Writer-Pro (Virtek Vision Inc., Waterloo, Ontario, Canada) microarrayer. The slides are blocked after spotting with 1% ammonia solution and stored at 4°C until needed. Washing steps with 95°C water for 2 min, 10 min in 100 mM NaOH and three times for 2 min in 95°C ddH2O were performed before APEX reactions to reduce the background fluorescence and remove unbound oligonucleotides.

The template DNA for APEX is amplified by PCR. We are using double-stranded DNA for APEX, which allows reading the sequence from both DNA strands simultaneously. A critical parameter in the APEX assay is the length of the target. We found that approximately 100 bp were optimal, PCR products of up to 200 bp could be used directly in the assay, but longer ones needed fragmentation before use in the APEX reaction. For this, a fraction of the dTTPs is replaced by

![FIGURE 13.1 Principle of the arrayed primer extension (APEX) approach. Courtesy of Asper Biotech Ltd., Tartu, Estonia.](image-url)
dUTPs in the amplification mix allowing for later treatment with Uracil N-Glycosylase (UNG)\textsuperscript{39} and heat treatment. The reaction mixture contained 1U thermolabile uracil-N-glycosylase (UNG) (Epicentre Technologies, Madison, WI) in UNG reaction buffer [final concentration 50 mM Tris-HCl, pH 9.0; 20 mM (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4}] incubated at 37°C for 1 h. After UNG treatment and heat treatment for 10 min at 95°C a fraction of reaction product could be controlled for the fragmentation efficiency by electrophoresis in a 2.5% agarose gel. No intact PCR product should be visible in the gel. By changing the dUTP:dTTP ratio, the mean length of the template DNA fragments can be varied.

Single nucleotide extension reactions can only work if no deoxyribonucleotide triphosphates are carried over from the amplification mix. A reliable method to inactivate dNTP left over from the PCR is enzymatic digestion with Shrimp Alkaline Phosphatase (SAP). This can be performed simultaneously with the UNG treatment and must be followed by thermal inactivation of both enzymes prior to the APEX reaction.

We used approximately 1.5 to 2 µg fragmented amplification product mix for the APEX assay. Engineered DNA polymerases\textsuperscript{30} are able to incorporate the dye terminators quite efficiently. The reaction mix contained 32 U of Thermo Sequenase DNA polymerase (Amersham Biosciences) Thermo Sequenase reaction buffer [final concentration 520 mM Tris-HCl, pH 9.5, 130 mM MgCl\textsubscript{2}] (Amersham Biosciences), and 12 µM final concentration of each fluorescently labeled ddNTPs: Texas Red-ddATP, Cy3-ddCTP, Fluorescein-ddGTP, Cy5-ddUTP (Amersham Biosciences, Perkin-Elmer Lifesciences, Boston, MA). The DNA in buffer was denatured at 95°C for 10 min. The enzyme and dye were immediately added to other components and the whole mix was applied to prewarmed slides at 58°C. The reactions were allowed to proceed 20 min under cover slips and stopped by washing at 95°C for two times 90 sec in ultrapure (MilliQ, Millipore Corp, Bedford, MA) water. Incorporation of labeled terminators is a very quick reaction, but hybridization is equilibrium process and thus it needs time to get strong enough signals. After reaction and washing, a droplet of SlowFade\textsuperscript{8} Light Antifade Reagent (Molecular Probes, Eugene, OR) was applied to the chips to limit bleaching of the fluorescein, covered with coverslip and imaged with Genorama\textsuperscript{TM} imaging system, at 20 µm resolution. All genotypes were identified by Genorama\textsuperscript{TM} 3.0 genotyping software (Asper Biotech Ltd., www.asperbio.com). The signal intensities of greyscale pictures from different fluorochromes were first equalized. The strongest signal was the base called. If the next strongest signal from both strands had an intensity level higher than 30 to 50% of the strongest signal, the position was called heterozygous. The sequence was compared with a reference and diverging bases indicated. All divergences and heterozygous positions were verified manually by comparing signals from different images by eyesight and using histogram values.

![FIGURE 13.2 Four-channel gray scale images for each fluorescent terminator nucleotides are used for APEX analysis. Two signals from duplicate oligonucleotides screening the same position in the analyzed DNA, each from sense- and antisense-strand, are obtained. Two upper rows show a homozygote and the lower row shows a heterozygote genotype. Courtesy of Asper Biotech Ltd., Tartu, Estonia.](image-url)
V. DISCUSSION OF PROBLEMS AND PRACTICAL HINTS

A. PRACTICAL HINTS FOR THE PRIMER DESIGN

We use the PRIMER3 program routinely with the following modified parameters:

- PRIMER_PRODUCT_OPT_SIZE = 200
- PRIMER_PRODUCT_SIZE_RANGE = 100–600
- PRIMER_OPT_SIZE = 21
- PRIMER_MIN_SIZE = 18
- PRIMER_MAX_SIZE = 26
- PRIMER_OPT_TM = 62
- PRIMER_MIN_TM = 59
- PRIMER_MAX_TM = 65
- PRIMER_MAX_DIFF_TM = 4
- PRIMER_OPT_GC_PERCENT = 35
- PRIMER_MIN_GC = 20
- PRIMER_MAX_GC = 60
- PRIMER_SALT_CONC = 20

Using a repeat library (PRIMER_MISPRIMING_LIBRARY) is strongly recommended. We suggest using a default library with the four most frequent repeat sequences, Alu, Line, Mir, and Mir2. PRIMER3 is relatively inefficient in doing searches against these repeats and adding all known repeats to this library would significantly slow down the primer design process. Masking template regions with low complexity with the DUST program\(^\text{31}\) might be a good idea. Using RepeatMasker\(^\text{32}\) is less practical, because it tends to mask long contiguous areas, even if the region contains nonrepeated segments, suitable for primer design. None of the currently available masking methods guarantees uniqueness of the designed PCR product, so the genome test is necessary in any case.

B. PRACTICAL HINTS FOR THE GENOME TEST

We suggest using BLAST,\(^\text{9}\) MEGABLAST,\(^\text{10}\) or SSAHA\(^\text{11}\) for the genome test. BLAST requires less memory than others if run on each chromosome separately (i.e., 400 MB for chromosome 1). MEGABLAST is faster than BLAST, but it does not handle assembled chromosome sequences properly, possibly because of long stretches of undefined nucleotides (N's) that are used to fill gaps in golden path. We had to create a full genomic database from all contigs and this required 1GB RAM for the MEGABLAST search. The locations on contigs were converted back to chromosome locations using a locally installed ENSEMBL database.\(^\text{12}\) SSAHA uses preindexed hash tables to look up the locations of strings in the genome. It requires more memory than other programs (about 1 GB RAM for larger chromosomes) but it is also significantly faster than others if the number of primer pairs in test exceeds 100.

The command line parameters for each program were:

- BLAST: formatdb -i chr1.fas -p F -A T
- BLAST: blatall -p blastn -i input.fas -d chr1.fas -F F
- MEGABLAST: megablast -i input.fas -d contigs.fas -F F -W 12 -D 2
- SSAHA INDEXES: ssaha chr1.fas -ph -sf fasta -sn chr1 -wl 10 -sl 1
- SSAHA SEARCH: ssaha input.fas chr1 -qd fasta -sf hash -pf -mp 15

BLAST and MEGABLAST should be run with the DUST filter turned off by defining `-F F', otherwise some of the primer binding sites might overlap masked areas and remain undetected. The search result of each program was parsed with specific parser written in PERL.
C. PRACTICAL HINTS FOR APEX GENOTYPING

APEX genotyping is a method with the smallest number of steps one can imagine and therefore there are few opportunities for mistakes. One is PCR product fragmentation. We test fragmentation on gel electrophoresis whenever we change the enzyme or protocol. Unfragmented PCR product as a template reduces the signal. The next step, primer extension, is Mg\(^{2+}\) sensitive. Thus, if the molar amount of the template is increased (large arrays with over 1000 extension reactions), the Mg\(^{2+}\) concentration has to be adjusted to a higher value.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

In brief, optimal primers for the genomic PCR are longer than 20 nucleotides and AT-rich. Primer binding sites should be rare in the genome and the generated product should be unique. The uniqueness of the product can be achieved by the genome test — counting of all possible binding sites and products in the genome for any given primer pair. We suggest the programs MEGABLAST,\(^{10}\) SSAHA,\(^{11}\) or GenomeTester\(^{13}\) for large-scale genome tests. On a smaller scale, BLAST\(^{9}\) is a slower but simpler alternative. BLAST is also able to report binding sites with mismatches, which slightly increases its sensitivity in the genome test. Unfortunately there is, at the moment, no method for the genome test which would permit appropriate modeling of primer binding sites. Nearest-neighbor algorithms\(^{33}\) would yield a higher sensitivity in the genome test, but are computationally extremely intensive and so far unpracticable.

Multiplex PCR is essential for large-scale projects as it helps to reduce the cost of PCR by an order of magnitude. Dedicated software is necessary to test all possible primer–primer interactions and to form interaction-free groups. A promising technical solution for more efficient multiplex PCR might be PCR suppression, which requires only one primer.\(^{34}\) As this technology requires 2 times less sequence-specific primers, there is also 4 times less chance for primer–primer interactions and higher multiplexing level could be achieved.\(^{35}\) Most of the current primer design methods use identity-based alignments to measure the strength of primer–primer interactions.

REFERENCES


29. Cronin, M. T., Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays, *Hum. Mutat.*, 7, 244, 1996.


14 Surface Plasmon Resonance Based Biosensor Technology for Real-Time Detection of PCR Products

Giordana Feriotto and Roberto Gambari

CONTENTS

I. Introduction..................................................................................................................141
II. Materials and Methods...............................................................................................143
   A. Primers, Molecular Probes, and Target Oligonucleotides........................................143
   B. DNA Templates and PCR Conditions .....................................................................143
   C. SPR-Based BIA..........................................................................................................143
      1. Protocol A: Determination of the Optimal Conditions for Hybridization,
         Including Detection of Point Mutations .................................................................144
      2. Protocol B: Immobilization of Target PCR Products and Injection
         of Molecular Probes................................................................................................145
      3. Protocol C: Immobilization of Molecular Probes and Injection of Target
         Asymmetric PCR Products .....................................................................................145
III. Results.......................................................................................................................146
   A. Choice of Primers and Probes ..................................................................................146
   B. Characterization of PCR Products Immobilized on Flow Cells...............................147
   C. Use of Peptide Nucleic Acids as Molecular Probes.....................................................149
   D. Injection of Asymmetric PCR Products to Flow Cells Carrying Specific Probes......149
IV. Discussion................................................................................................................149
   A. Advantages of SPR-Based BIA for Characterization of PCR Products......................149
   B. Problems and Possible Solutions.............................................................................151
   C. Future Perspectives..................................................................................................152
Acknowledgments .............................................................................................................152
References ............................................................................................................................152

I. INTRODUCTION

Surface Plasmon Resonance (SPR) and biosensor technologies for biospecific interaction analysis (BIA)\(^1\)\(^2\)\(^3\) enable the monitoring of DNA:DNA and DNA:RNA hybridization in real time\(^3\)\(^4\). In the BIAcore biosensor system (Figure 14.1A), plane polarized light is reflected from the gold-coated sensor chip (Figure 14.1B), where the molecular interactions take place\(^5\). Surface plasmon resonance in the gold layer results in the extinction of the reflected light at a specific angle. This angle (SPR angle) varies with the refractive index of the solution close to the other side of the sensor chip.
When molecules (ligands) bind to the chip, the refractive index changes, and with it the SPR angle. The change of the latter is measured in resonance units (RU). After ligand immobilization, the injection of analyte(s) results in a further increase of RU only when molecular interactions between the ligand and the analyte occur. Moreover, injection of binding/running buffer allows a determination of whether this interaction is stable or not. Sensor chips can be regenerated by removing all the bound analyte by short pulses with suitable buffers.

The change in refractive index is proportional to the change in absorbed mass, permitting the quantitative monitoring of ligand interaction, such as nucleotide hybridization, in real time. Therefore, SPR-based BIA has been proposed as a tool for detecting and characterizing PCR products in a number of biological applications, including detection of HIV-1 infection, genetic mutations (ΔF508 and W1282X) involved in cystic fibrosis (CF), and Roundup-Ready™ soybean sequences. The general setup of SPR-based BIA for DNA:DNA hybridization is shown in Figure 14.1. Biotinylated target oligonucleotides, complementary to the nucleotide sequences to be studied, are immobilized on streptavidin-coated flow cells (sensor chip SA) and molecular probes injected. For SPR-based characterization of PCR products of diagnostic relevance, two different approaches are usually considered. In the first, target PCR products are immobilized on the chip, and the probe (oligodeoxyribonucleotide [ODN], peptide nucleic acid [PNA], RNA, or PCR products) is injected. In the second, molecular probes (usually ODN, PNA, or PCR products) are immobilized on the chip, injecting the PCR products to be analyzed.
Following the injection of ligand, the sensorgram reproducibly shows a gradual increase of RU (starting from initial RU, $RU_i$) up to the final RU values ($RU_{\text{fin}}$) at the end of the injection. The amount of bound probe can be therefore calculated by subtracting the $RU_i$ values from the $RU_{\text{fin}}$ values ($RU_{\text{fin}} - RU_i$). After washing with HEPES-buffered Saline-EP (HBS-EP), the RU may drop or remain constant, depending on the stability of the DNA:probe complex. The RU remaining after this step is referred to as $RU_{\text{res}}$. In some cases, the informative values are $RU_{\text{res}} - RU_i$.

With respect to molecular probes, recent published reports indicate that peptide nucleic acids (PNAs)\textsuperscript{12} could be of great interest.\textsuperscript{13–16} PNAs are indeed DNA mimics able to hybridize with complementary DNA with high efficiency, since in these molecules the negatively charged sugar-phosphate backbone is replaced by neutral $N$-(2-aminoethyl)glycine units.\textsuperscript{12} Since no electrostatic repulsion occurs during DNA:PNA hybrids formation, even short PNAs are expected to hybridize efficiently to target DNA.\textsuperscript{14} In addition, unlike oligonucleotides, PNAs are expected to bind with high efficiency to single stranded PCR products since their binding could be independent from the secondary structure of target DNA.\textsuperscript{14} Nevertheless, the stability of PNA:DNA hybrids is greatly affected by the presence of a single base mismatch.\textsuperscript{14,15}

II. MATERIALS AND METHODS

A. PRIMERS, MOLECULAR PROBES, AND TARGET Oligonucleotides

Synthetic HPLC-purified oligonucleotides were purchased from a variety of molecular biology companies such as Sigma-Genosys (Cambridge, U.K.). PNAs were obtained from Professor Rosangela Marchelli (Department of Organic and Industrial Chemistry, Parma University, Italy), but can also be purchased from several biotechnology companies such as PRIMM (S. Raffaele Biomedical Science Park, Milan, Italy). All primers, ODN and PNA probes, and biotinylated target oligonucleotides were resuspended in sterile distilled water to obtain a 100 $\mu$M stock solution.\textsuperscript{5,10}

B. DNA Templates and PCR Conditions

Genomic DNA was prepared by lysing white blood cells with 600 $\mu$g/ml of Proteinase K (Eurobio, Les Ulis Cedex, France) in 50 mM Tris-HCl pH 8, 100 mM EDTA, 100 mM NaCl, 1% SDS for 12 h at 55°C. DNA was then purified by phenol extraction and ethanol precipitation using standard methods.\textsuperscript{10} For extraction of soybean DNA, the Wizard Magnetic DNA Purification System for food (Promega Corporation, Madison, WI) was used.\textsuperscript{11} Sequences of PCR primers and DNA or PNA probes used in the experiments reviewed in this work for identification of cystic fibrosis W1282X mutation,\textsuperscript{9,10} gag HIV-1 sequences,\textsuperscript{5,7} and Roundup-Ready soybean sequences\textsuperscript{11} are shown in Table 14.1. PCR experiments were performed using a Perkin–Elmer thermal cycler (GeneAmp PCR System 9600). Each PCR reaction was performed in a final volume of 50 to 100 $\mu$l, containing 50 to 500 ng of target genomic DNA, 50 mM KCl, 10 mM TRIS-HCl pH 8.8, 1.5 mM MgCl$_2$, 33 $\mu$M dNTPs, 0.33 $\mu$M PCR primers, and 2 U/reaction of Taq DNA polymerase (Finnzymes Oy, Espoo, Finland). We usually performed 30 to 40 PCR cycles; conditions for the denaturation steps were 30 sec, 95°C; specific experimental conditions for annealing and extension are reported in Table 14.1. Biotinylated PCR products were generated using an excess (two-fold) of the unbiotinylated primer. Asymmetric PCR products were obtained using a 10 to 100:1 primer ratio.\textsuperscript{7–11}

C. SPR-BASED BIA

BIACore 1000\textsuperscript{TM} analytical system (BIACore AB, Uppsala, Sweden) was used in all experiments. Sensor chips SA, precoated with streptavidin, and the running buffer HEPES buffered saline-EP (HBS-EP), containing 10 mM HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% (v/v) Surfactant P20, were from BIACore AB (Uppsala, Sweden).

In protocol A, a target oligonucleotide is immobilized on a flow cell of the sensor chip, and molecular probes are injected. This protocol is routinely employed to determine experimental conditions for optimal hybridization. When a target biotinylated oligonucleotide is immobilized on a SA sensor-chip, the binding kinetic is fast and DNA remains bound, even after NaOH washing. When designing molecular probes for the detection of point mutations, great attention should be paid to the length of synthetic oligonucleotides employed. Under standard BIA conditions (HBS-EP, 25°C), oligonucleotides longer than 14 bp usually do not discriminate efficiently between full-matched and mismatched target biotinylated oligonucleotides immobilized onto the sensor chip.

---

**TABLE 14.1**

**SPR-Based BIA of PCR Products: Sequences of Used ODN and PNA Molecules**

<table>
<thead>
<tr>
<th>Detection of CF W1282X Mutation</th>
<th>CF3 (biotinylated PCR primer)</th>
<th>5'-AAGGAGAAATCCAGATCGA-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CF2 (PCR primer)</td>
<td>5'-GCTCACCTGTGGTATCCT-3'</td>
</tr>
<tr>
<td></td>
<td>M-CF-21 (mutated biotinylated target DNA)</td>
<td>5'-Biot-TGCAACAGTGAAGGAAAGCT-3'</td>
</tr>
<tr>
<td></td>
<td>N-CF-17 (DNA probe)</td>
<td>5'-GCTTTCCTCCACTGTTTG-3'</td>
</tr>
<tr>
<td></td>
<td>N-CF-12 (DNA probe)</td>
<td>5'-TTTCTTCCACTG-3'</td>
</tr>
<tr>
<td></td>
<td>N-CF-9 (DNA probe)</td>
<td>5'-TCCTTCCACT-3'</td>
</tr>
<tr>
<td></td>
<td>M-CF-17 (DNA probe)</td>
<td>5'-GCTTTCCTCCACTGTTTG-3'</td>
</tr>
<tr>
<td></td>
<td>M-CF-12 (DNA probe)</td>
<td>5'-TTTCTTCCACTG-3'</td>
</tr>
<tr>
<td></td>
<td>M-CF-9 (DNA probe)</td>
<td>5'-TCCTTCCACT-3'</td>
</tr>
<tr>
<td></td>
<td>N-PNA-CF-9 (PNA probe)</td>
<td>H-TCTTCACAT-NH2</td>
</tr>
<tr>
<td></td>
<td>M-PNA-CF-9 (PNA probe)</td>
<td>H-TCTTCACAT-NH2</td>
</tr>
<tr>
<td>PCR conditions:</td>
<td>annealing, 30 sec, 58°C; elongation, 20 sec, 72°C</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Detection of HIV1 Gene Sequences</th>
<th>SK145 (PCR primer)</th>
<th>5’-AGTGGGGGGACATCAAGCAGCCATGCAAAT-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SK39 (PCR primer)</td>
<td>5’-TTTGGTCCTTGTCTTATGTCCAGAATGC-3’</td>
</tr>
<tr>
<td></td>
<td>SK150 (DNA probe)</td>
<td>5’-TGCTATGTCACCTCCCCCTTGTTCTTC-3’</td>
</tr>
<tr>
<td>PCR conditions:</td>
<td>annealing, 1 min 55°C; elongation, 1 min, 72°C</td>
<td></td>
</tr>
<tr>
<td>Asymmetric PCR:</td>
<td>SK145:SK39 (10:1) 250 ng/25 ng</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Detection of Roundup Ready Soybean Transgene</th>
<th>RupR1 (PCR primer)</th>
<th>5’-TGTAATCCCTTGACACGCGATGTTG-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RupR2 (PCR primer)</td>
<td>5’-CGCACAATCCCACTCGTTCTC-3’</td>
</tr>
<tr>
<td></td>
<td>biot-RupR2 (biotinylated PCR primer)</td>
<td>5’-biot-CGCAACATCCCACTCGTTCTC-3’</td>
</tr>
<tr>
<td></td>
<td>Lec1 (PCR primer)</td>
<td>5’-ATGGGGCTTGCCCTTTTTTC-3’</td>
</tr>
<tr>
<td></td>
<td>Lec2 (PCR primer)</td>
<td>5’-CGAAACCCCTCAGAGGATGGA-3’</td>
</tr>
<tr>
<td></td>
<td>biot-Lec2 (biotinylated PCR primer)</td>
<td>5’-biot-CGAAACCCCTCAGAGGATGGA-3’</td>
</tr>
<tr>
<td></td>
<td>RupR-15 (DNA probe)</td>
<td>5’-CTAGAGTCAAGCTTTG-3’</td>
</tr>
<tr>
<td></td>
<td>Lec-15 (DNA probe)</td>
<td>5’-TCAAGTCGCTGCTGTTG-3’</td>
</tr>
<tr>
<td>PCR conditions:</td>
<td>annealing, 30 sec, 62°C; elongation, 10 sec, 72°C</td>
<td></td>
</tr>
<tr>
<td>Asymmetric PCR:</td>
<td>RupR1:RupR2 or Lec1:Lec2 (50:1) 300 ng/16 ng</td>
<td></td>
</tr>
</tbody>
</table>
2. Protocol B: Immobilization of Target PCR Products and Injection of Molecular Probes

The first step of this protocol is to produce a biotinylated double-stranded PCR product. In general, it is advisable to employ an excess of nonbiotinylated over biotinylated primer for the production of target PCR products. This minimizes the amount of unincorporated biotinylated primer in the product mixture. In the second step of this protocol, the final PCR product is further purified by Microcon-30 (Millipore Corporation, Bedford, MA). The third step is the injection of the biotinylated PCR product on flow cells of a SA sensor chip. In this case, the binding kinetic is slow (Figure 14.2) and repeated injections of biotinylated PCR products are usually necessary to reach saturation. The regeneration step with 50 mM NaOH induces a decrease of the RU to about 1/2 of the RU_{bound}. This is the RU_{rest}, caused by single stranded (ss) PCR product stably immobilized on the sensor chip by biotin–streptavidin interactions. After these steps, molecular probes are injected.

3. Protocol C: Immobilization of Molecular Probes and Injection of Target Asymmetric PCR Products

The first step of this protocol employs either biotinylated ODNs, PNAs, or PCR products to be used as molecular probes immobilized on sensor chip flow cells. In this protocol the target DNA molecules to be analyzed is usually produced by asymmetric PCR.

Table 14.2 summarizes reports present in the recent literature on the use of the described SPR-BIA protocols for detection and characterization of PCR products.
III. RESULTS

A. CHOICE OF PRIMERS AND PROBES

In SPR-based analysis of PCR products, the choice of the primers is relevant, at least for some of the experimental approaches described. In protocol C, immobilized probes are employed and asymmetric PCR products are injected. Therefore, primers should be designed and synthesized for optimal yields of single-stranded asymmetric PCR products. In our hands, this is an important parameter. In addition, single-stranded PCR products (either injected as for protocol C or immobilized on the chip as for protocol B) should exhibit low levels of secondary structure in the BIA experimental conditions, in order to maximize the hybridization efficiency with molecular probes. Accordingly, it is advisable to study possible secondary structures with the help of dedicated software. For instance, the mfold software (version 3.0) developed by Zuker et al.\textsuperscript{20} and Mathews et al.\textsuperscript{21} can determine secondary structures of single-stranded PCR products. The analysis should be performed with parameters identical to the BIA experiment, such as the standard conditions 25°C and 0.15 M NaCl.

The length of the probes is also very important for the detection of point mutations with BIA. Figure 14.3A shows representative results obtained after injection of a 15-mer probe to a target complementary biotinylated oligonucleotide immobilized on the sensor chip. RU\textsubscript{i}, RU\textsubscript{fin}, and RU\textsubscript{res} are the resonance units before injection (RU\textsubscript{i}), at the end of the injection (RU\textsubscript{fin}), and after washing with hybridization buffer (RU\textsubscript{res}). Figure 14.3B demonstrates that both N-CF-17 and M-CF-17 ODN probes hybridize to the target M-CF-21 DNA generating stable DNA:DNA hybrids. Therefore, shorter oligonucleotides should be used for diagnostic purposes. For instance, while both N-CF-12 and M-CF-12 probes hybridize to the target M-CF-21 DNA, the generated DNA:DNA hybrids exhibit different stability. Therefore, point mutations can be identified by looking at the difference between residual and initial resonance units (RU\textsubscript{res} – RU\textsubscript{i}) (Figure 14.3C). Finally, shorter oligonucleotide probes (N-CF-9 and M-CF-9) allow the identification of the W1282X point mutation during the association phase; in this case the difference between final and initial resonance units (RU\textsubscript{fin} – RU\textsubscript{i}) is informative (Figure 14.3D). While the results shown in panels C and D of Figure 14.3 are both informative, the use of longer probes (such as N-CF-12 and M-CF-12 probes) generates higher RU levels and, therefore, higher sensitivity. This is a very important parameter when these probes are used on immobilized PCR products.

---

**TABLE 14.2**
Examples of SPR-Based Biosensor Technology for Real-Time Detection of Polymerase-Chain Reaction (PCR) Products

<table>
<thead>
<tr>
<th>Organism</th>
<th>Gene/Mutation</th>
<th>Probe</th>
<th>Protocol\textsuperscript{a}</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV-1</td>
<td>Verotoxin 2 subunit A</td>
<td>ODNs</td>
<td>C</td>
<td>5</td>
</tr>
<tr>
<td>E. coli</td>
<td>p53 gene mutations</td>
<td>ODNs</td>
<td>C</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>CF ΔF508 mutation</td>
<td>ODNs</td>
<td>B, C</td>
<td>6</td>
</tr>
<tr>
<td>Human</td>
<td>CF W1282X mutation</td>
<td>ODNs</td>
<td>A, C</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>PNA</td>
<td>PNAs</td>
<td>A, B</td>
<td>9</td>
</tr>
<tr>
<td>GMO,\textsuperscript{b} Soybean</td>
<td>RoundUp Ready</td>
<td>ODNs</td>
<td>A, B, C</td>
<td>10</td>
</tr>
<tr>
<td>GMO,\textsuperscript{b} common markers</td>
<td>P35S promoter, terminator TNOS</td>
<td>ODNs</td>
<td>A, C</td>
<td>11</td>
</tr>
</tbody>
</table>

\textsuperscript{a} See text for details.

\textsuperscript{b} Genetically-modified organism.
B. CHARACTERIZATION OF PCR PRODUCTS IMMOBILIZED ON FLOW CELLS

In a typical experiment, PCR products are immobilized and molecular probes are injected in order to detect the presence or absence of a sequence of interest. This approach has been applied to the detection of Roundup-Ready and lectin soybean gene sequences by injecting RupR-15 or Lec-15 probes (see Table 14.1) to flow cells carrying Roundup-Ready or lectin PCR products. These oligonucleotides generate stable hybrids with the target sequence and high RU\textsubscript{fin} values.\textsuperscript{11} No cross hybridization of Roundup-Ready probes with immobilized lectin PCR products, or of lectin probes with immobilized Roundup-Ready PCR products is detectable.\textsuperscript{11}

Another example, the identification of the CF-associated point mutation W1282X, is shown in Figure 14.4. Here, we employed immobilized PCR products from normal W1282X heterozygous and W1282X homozygous samples. We found that 12-mers are useful to detect W1282X CF mutations when stability of the generated hybrids is analyzed (RU\textsubscript{res} – RU\textsubscript{i} values). We determined...
the cystic fibrosis index (CF index) as the value \((RU_{res} - RU_{i})(N)/(RU_{res} - RU_{i})(M)\), where \((RU_{res} - RU_{i})(N)\) are the values obtained with the “normal” N-CF-12 ODN probe and the \((RU_{res} - RU_{i})(M)\) values are those obtained with the “mutated” M-CF-12 ODN probe. The CF index was found high (4.3 ± 0.8) when PCR products from normal subjects were employed (see the representative sensorgrams shown in Figure 14.4A). On the contrary, this value approached 1 (1.05 ± 0.35) when
PCR products from heterozygous W1282X subjects were employed (see Figure 14.4B). Finally, the CF index was found always lower than 0.5 (0.28 ± 0.1) when PCR products from homozygous W1282X samples were immobilized on the SA sensor chip (see Figure 14.4C). Thus, SPR-based BIA allows the detection of point mutations and discrimination between heterozygous and homozygous carriers.

C. USE OF PEPTIDE NUCLEIC ACIDS AS MOLECULAR PROBES

In the experiment shown in Figure 14.5 the hybridization behavior of 9-mer PNA probes was compared to that of 9-mer ODN probes. Interestingly, both the N-PNA-CF-9 and M-PNA-CF-9 hybridize to the relative full complementary target DNAs (Figure 14.5 and data not shown). In this case, hybridization is much more efficient than that of the 9-mer ODN probes and the generated PNA:DNA hybrids are much more stable than DNA/DNA hybrids. No hybridization occurs between mismatched PNA probes and target W1282X DNA (Figure 14.5 and data not shown). When PNA probes were injected to immobilized PCR products and compared to the hybridization behavior of 9-mer ODN probes, it was found that only PNA are able to give informative results.

D. INJECTION OF ASYMMETRIC PCR PRODUCTS TO FLOW CELLS CARRYING SPECIFIC PROBES

This protocol was demonstrated to be feasible for detection of specific gene sequences, as in the case of HIV-1 infection, and in a recently reported study on identification of genetically modified organisms (GMO). In the case of HIV-1 detection, gag-specific probes recognizing a region of the HIV-1 genome were immobilized. The data obtained demonstrate that this technology is suitable for identifying HIV-1 sequences after direct injection of either purified asymmetric PCR products or unpurified asymmetric PCRs. Table 14.3 shows a comparison of the hybridization efficiency when flow cells carrying immobilized ODN and PCR product probes are used. These data were obtained in a study of soybean Roundup-Ready GMO sequences. The results obtained with asymmetric Roundup Ready and lectin PCR product injected into flow cells carrying either ODN or PCR product probes demonstrate that, although the values of RU$_{in}$ – RU$_i$ were significantly different, both approaches lead to informative results.

IV. DISCUSSION

A. ADVANTAGES OF SPR-BASED BIA FOR CHARACTERIZATION OF PCR PRODUCTS

The major advantages of using SPR-based analysis employing optical-based biosensors are (1) analysis is performed in few minutes, (2) no label is required, and (3) in protocols employing immobilization of molecular probes on the flow cells, the sensor chip can be reused several times. The main limit is the cost of most of the commercially available instruments.

With respect to the use of (RU$_{in}$ – RU$_i$) or (RU$_{res}$ – RU$_i$) values, the choice depends on the ability of the employed probes in discriminating target DNA during the association phase of the BIA analysis. In this case, the higher (RU$_{in}$ – RU$_i$) values are preferably used for analytical determinations, with highly reproducible results (for instance in the detection of GMO or HIV-1 sequences). We cannot generalize the use of (RU$_{in}$ – RU$_i$) values for the detection of point mutations; this strongly depends on the length of the probes and the secondary structures of PCR-generated single-stranded target DNAs. In most instances, when (RU$_{in}$ – RU$_i$) values are not informative for detecting point mutations, (RU$_{res}$ – RU$_i$) values should be taken into consideration (such in the case shown in Figure 14.4). However, we recently demonstrated that (RU$_{in}$ – RU$_i$) differences were informative for detecting the $\beta^+$-thalassemia mutation (Feriotto et al., manuscript in preparation).
FIGURE 14.5 Sensograms obtained after injection of 25 µl containing 0.5 µg of normal (dotted lines) and mutated (solid line) W1282X CF 9-mer probes to SA sensor chip flow cells carrying normal W1282X 12-mer target DNA. (A) PNA, (B) ODN, both dissolved in HBS-EP. (Modified from Ferioto, G. et al., Peptide nucleic acids and biosensor technology for real-time detection of the cystic fibrosis W1282X mutation by surface plasmon resonance, Lab. Invest., 81, 1415, 2001.)
A partial list of problems and possible solutions is shown in Table 14.4. In order to maximize the hybridization efficiency in SPR-based analysis of PCR products, the choice of the probes is crucial. It appears to be important to determine possible secondary structures of target PCR products; in fact, probes hybridizing with single-stranded PCR regions not involved in secondary structures are expected to be more efficient than probes recognizing double stranded stretches of target PCR product. The choice of the PCR primers as well as the possible use of PNAs is relevant, at least for some of the experimental approaches described. In addition, when

### TABLE 14.4
Problems and Possible Solutions of SPR-Based BIA of PCR Products

<table>
<thead>
<tr>
<th>Problems</th>
<th>Possible Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymmetric PCR is not produced efficiently</td>
<td>Synthesize two asymmetric but complementary PCR</td>
</tr>
<tr>
<td></td>
<td>Products differing in length and exposing after annealing the sequence recognized by probes</td>
</tr>
<tr>
<td></td>
<td>Produce ds PCR products using a biotinylated PCR primer and purify single stranded target molecules</td>
</tr>
<tr>
<td></td>
<td>Produce ds PCR products using a chimeric 5’-RNA-DNA-3’ primer followed by degradation of the 5’ RNA stretch</td>
</tr>
<tr>
<td>Heavy secondary structure of asymmetric PCR products</td>
<td>Change PCR primers</td>
</tr>
<tr>
<td></td>
<td>Increase hybridization temperature</td>
</tr>
<tr>
<td></td>
<td>Change hybridization buffers</td>
</tr>
<tr>
<td></td>
<td>Use PNA probes</td>
</tr>
<tr>
<td></td>
<td>Synthesize two asymmetric but complementary PCR products differing in length and exposing after annealing the sequence recognized by probes</td>
</tr>
<tr>
<td></td>
<td>Produce ds PCR products using a chimeric 5’-RNA-DNA-3’ primer followed by degradation of the 5’ RNA stretch</td>
</tr>
<tr>
<td>Hybridization to immobilized probes is not efficient</td>
<td>Change PCR primers in order to modify possible secondary structure of asymmetric PCR product</td>
</tr>
<tr>
<td>Probes do not discriminate between full-matched and mismatched target DNA</td>
<td>Increase the distance between biotin and probe sequences</td>
</tr>
<tr>
<td>High bulk effects of PCR products</td>
<td>Purify PCR products with Microcon-30</td>
</tr>
</tbody>
</table>

www.taq.ir
secondary structure is a problem, one of the possible solutions is to follow the approach described by Kai et al.,\textsuperscript{17} producing two asymmetric, but complementary, PCR products of different length. After heat denaturation and gradual cooling, these anneal to generate double-stranded molecules with protruding ssDNA ends exposing the probe-binding site.\textsuperscript{17} If the yields of asymmetric PCR are low, a biotinylated primer and streptavidin-coated magnetic beads can be helpful to capture the PCR product strands.\textsuperscript{19}

\section*{C. Future Perspectives}

In our opinion, two major issues should be investigated in the near future. These are the possibility to perform multiplex analysis and the suitability of this technology to quantify the proportion of a given gene sequence within a sample. The first issue is important for diagnosing complex genetic mutations; the latter is of great interest in applications such as quantifying the presence of GMO in foods.

We would like to make a final comment on recently published data describing SPR array biosensors based on multielement transduction systems combined with conventional SPR spectroscopy for monitoring binding events on macro- or micro-patterned arrays created on disposable sensor chips.\textsuperscript{23} These effective platforms allow the analysis of several independent biospecific binding events simultaneously.\textsuperscript{24} Studies aimed at developing SPR-based arrayed biosensors for applications in the field of DNA:DNA hybridization are expected to generate very interesting results in the near future.\textsuperscript{24,25}

\section*{Acknowledgments}

This research was supported by Ministero della Sanità, Italy, by P. F. Biotecnologie and by MURST Cofin-2000 (Tracciabilità di frammenti di DNA e proteine lungo la filiera alimentare e mangimistica con metodi applicati alla etichettatura volontaria — EC1139/98).

\section*{References}

15 Fluorescent Amplified Fragment Length Polymorphism (FAFLP) Genotyping

Catherine Arnold

CONTENTS

I. Introduction ................................................................................................................................................. 155
   A. Background ........................................................................................................................................ 155
   B. FAFLP .............................................................................................................................................. 156
II. Materials and Methods ............................................................................................................................... 157
   A. Equipment ......................................................................................................................................... 157
   B. Materials .......................................................................................................................................... 157
   C. Method ............................................................................................................................................. 157
      1. Genomic DNA Digestion .................................................................................................................. 157
      2. Preannealed Adapter Ligation ........................................................................................................... 158
      3. PCR ................................................................................................................................................ 158
      4. Fragment Separation ......................................................................................................................... 158
III. Results ...................................................................................................................................................... 158
IV. Discussion ............................................................................................................................................... 160
Acknowledgment .......................................................................................................................................... 161
References ...................................................................................................................................................... 161

I. INTRODUCTION

A. BACKGROUND

The ultimate bacterial genotyping method must involve comparison of whole genome sequence data. Sequence data is now available for many bacterial genomes and, in some cases, for different isolates of the same species in various genome databases (e.g., www.tigr.org/tdb/mdb/mdbcomplete.html and www.ncbi.nlm.nih.gov). However, until the comparison of sequence data from every bacterial genome is technically and financially feasible, genome sampling will remain the method of choice. This sampling is usually in the form of some kind of restriction fragment length polymorphism (RFLP) analysis or sequencing of selected regions of the genome (multilocus sequence typing, MLST).¹ One of these genome-sampling methods, FAFLP, has emerged as the method of choice for a variety of applications, including outbreak investigation and molecular epidemiological studies.²-⁶
B. FAFLP

Based on the radioactive AFLP method described by Vos et al., the FAFLP method is a simple procedure that uses a specific PCR to select, amplify, and fluorescently label a subset of fragments generated by the restriction enzyme digestion of genomic DNA. These fragments are then detected and sized on an automated sequencer. The advantages of FAFLP include speed (compared with techniques such as pulsed-field gel electrophoresis [PFGE] and objectivity, as the fragments are precisely sized by machine against an internal size standard run together with the sample. As well as being precise, FAFLP data can also be checked against the sizes of FAFLP fragments predicted in silico using the complete genome of a sequenced bacterial strain of the same species. In silico predictions and in vitro FAFLP data compared surprisingly well (97 to 98%). Importantly, the fragments generated are derived from multiple sites throughout the genome, the results are not weighted in favor of particular genes such as those that code for an easily detectable phenotype (e.g., antibiotic resistance) or, as is the case for MLST, in favor of other selected genes (both “housekeeping” and more variable genes). Furthermore, FAFLP results have been shown to be reproducible between laboratories as long as generated with the same sequencer, the same run conditions, and therefore identical sizing.

The FAFLP method is performed as follows: DNA is extracted from a pure plating of a bacterial isolate. It is cut with two restriction enzymes, and double-stranded adapter oligonucleotides that will bind the primers of the PCR are then ligated to the digested fragments. Following PCR amplification in which one of the primers complementary to the adapter sequences is fluorescently labeled, the fragments are separated in the slab or capillary polyacrylamide gel/matrix of an automated sequencer. Each sample is run with a differently labeled size standard which, following automatic calculation and comparison with a standard curve, allows precise and reproducible sizing (±0.5 base pairs) of the FAFLP fragments. Only a proportion of the fragments are amplified. During the DNA digestion with two enzymes, hundreds or even thousands of fragments will be generated, many with the same restriction site at each end. However, although all of the fragments generated will be ligated to adapters, fragments with a different restriction site at each end are preferentially amplified and detected on the gel. The reason for this may be twofold. First, each strand of a fragment cut at both ends by the same enzyme will, following denaturation, form a stable “panhandle” structure, as their 5′ and 3′ ends are complementary because of the adapters being the same at both ends of the fragment. This effectively removes this group of molecules as target for the PCR. The degree to which this occurs will depend on whether intramolecular reactions (panhandle formation) are favored over intermolecular reaction (primer annealing), which will depend on the ratio of primer to PCR product concentration and annealing temperature. Second, only one primer is labeled, so only fragments cut with the corresponding restriction enzyme will be detected. PCR of fragments cut at both ends with the restriction enzyme, to which the unlabeled primer/adapter combination corresponds, will not be detected. Both reasons seem to explain why only the fragments cut with both enzymes should be detected. Another advantage of FAFLP is that its level of resolution (e.g., the number of useful data points or fragments generated) can be increased or decreased at will, depending on the application. Different enzymes will generate different numbers of fragments and, once an optimum enzyme combination is chosen, “selective” nucleotides can be added to the 3′ end of the PCR primer to limit the number of fragments that the primer will bind to. If, for example, an “A” selective nucleotide is added to the 3′ end of one of the primers, only fragments containing a complementary “T” immediately adjacent to the adapter/restriction site will be amplified, reducing the number of fragments generated to 1/4 of the original “nonselective” total (if the A:G:C:T ratio was the same). If another selective base were added to the same primer, the number of fragments would be reduced to 1/16th (i.e., 1/4 × 1/4) of the original number of fragments.
An optimized protocol for FAFLP analysis is given below. The enzyme combination most extensively used in our laboratory (EcoRI and MseI) has been described, but other combinations can be used depending on the organism being studied and the application.

**A. Equipment**

1. Automated sequencer equipped with fragment analysis software (including both slab gel and capillary platforms).
2. Thermal cycler capable of “touchdown” PCR (i.e., the ability to decrease the annealing temperature progressively with each cycle).

**B. Materials**

1. Restriction enzymes (e.g., EcoRI and MseI) and T4 DNA ligase (New England Biolabs [NEB], Hitchin, Hertfordshire, U.K.) and their appropriate buffers.
2. Preannealed oligonucleotide adapters specific to chosen restriction enzymes, and PCR primers (Table 15.1, taken from Vos et al.). One of these primers is fluorescently labeled. This is usually the primer that is directed to the adapter corresponding to the less frequently cutting enzyme, e.g., EcoRI. The differing amounts of primer added reflect this (see text below).
3. PCR reagents (dNTPs, Taq Polymerase, buffers, and MgCl2).
4. Size standard appropriate to the sequencer used, and deionized formamide.

**C. Method**

1. Genomic DNA Digestion

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 ng genomic DNA, dH₂O ad</td>
<td>20.0 μl</td>
</tr>
<tr>
<td>10 X MseI buffer (NEB)</td>
<td>2.0 μl</td>
</tr>
<tr>
<td>5 U MseI (NEB)</td>
<td>0.5 μl</td>
</tr>
<tr>
<td>10 μg/μl bovine serum albumen (NEB)</td>
<td>0.2 μl</td>
</tr>
<tr>
<td>DNase-free RNase A (10 μg/μl)</td>
<td>1.0 μl</td>
</tr>
</tbody>
</table>

**TABLE 15.1**

Sequences of Oligonucleotides Used as Primers

<table>
<thead>
<tr>
<th>Name</th>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoRI adapter 1</td>
<td>5’-CTCGTAGACTGCGTACC-3'</td>
</tr>
<tr>
<td>EcoRI adapter 2</td>
<td>3’-CTGACGCCATGGTTAA-5'</td>
</tr>
<tr>
<td>MseI adapter 1</td>
<td>5’-GACGATGAGTCCTGAG-3C</td>
</tr>
<tr>
<td>MseI adapter 2</td>
<td>3’-CTACTCAGGACTCAT-5'</td>
</tr>
<tr>
<td>EcoRI primer</td>
<td>5’-6FAM-GACTCGGTACCAATTC-3'</td>
</tr>
<tr>
<td>MseI primer</td>
<td>5’-GATGAGTCCGAGTAATA-3'</td>
</tr>
</tbody>
</table>

Incubate for 1 h at 37°C.

To this digest add:
- 5 U of EcoRI (Life Technologies, Paisley, U.K.) 0.5 µl
- 0.5 M Tris-HCl (pH 7.6) 1.7 µl
- 0.5 M NaCl 2.1 µl
- Total volume 22.0 µl
dH₂O 0.7 µl

Incubate the reaction mixture for a further hour at 37°C.

Inactivate endonucleases in the reaction prior to ligation by heating to 65°C for 10 min.

2. Preannealed Adapter Ligation

DNA, digested in step 1 25.0 µl
EcoRI adapter (2µM) 5.0 µl
MseI adapter (20µM) 5.0 µl
10 X ligase buffer (NEB) 5.0 µl
T4 DNA ligase (40 U/µl) 0.1 µl
dH₂O 9.9 µl
- Total volume 50.0 µl

Incubate the reaction mixture overnight at 12°C.

Inactivate by incubation at 65°C for 10 min.

3. PCR

Digestion/ligation reaction mixture 2.0 µl
10 X PCR buffer (Life Technologies, Paisley, U.K.) 2.0 µl
MgCl₂ (50 mM, Life Technologies) 1.0 µl
dNTPs (10 mM, Life Technologies) 0.4 µl
Fluorescently labeled primer (e.g., EcoRI primer, 10 µM) 0.4 µl
Unlabeled primer (e.g., MseI primer, 10 µM) 2.0 µl
0.5 U Taq polymerase (Life Technologies) 0.1 µl
dH₂O 12.1 µl
- Total volume 20.0 µl

Perform “touchdown” PCR: Denaturation for 2 min at 94°C (one cycle), followed by 30 cycles of denaturation at 94°C for 20 sec, a 30 sec annealing step (see text below), and a 2-min extension step at 72°C. The annealing temperature for the first cycle is 66°C; for the next nine cycles the temperature is decreased by 1°C at each cycle. The annealing temperature for the remaining 20 cycles is 56°C.

4. Fragment Separation

For the ABI Prism 377 DNA automated sequencer (Applied Biosystems), amplification products are separated on a 5% denaturing polyacrylamide gel. The sample mixture (1.0 µl of sample, 1.25 µl of deionized formamide, 0.25 µl dextran blue in EDTA, and 0.5 µl of internal size marker) is heated at 95°C for 2 min, cooled on ice, and immediately loaded onto the gel. The running buffer is 1 X TBE, and the electrophoresis conditions are 2 kV at 50°C for 14 h. The well-to-read distance is 48 cm.

III. RESULTS

An example of the use of the above protocol was for a study of Neisseria gonorrhoeae isolates from a U.K. town with known epidemiological features (Figure 15.1). The figure shows FAFLP
trace data for five \textit{N. gonorrhoeae} isolates, consisting of two epidemiologically linked pairs and an unlinked isolate. The boxed numbers below peaks indicate the size of each fragment in base pairs. Each pair of epidemiologically linked isolates share identical profiles and both of these profiles differ from the profile of the unlinked isolate. These data can also be presented as a binary matrix (Figure 15.2), which can then be further analyzed to create dendrograms if required.\textsuperscript{3-5,8,9}

\begin{table}[h]
\centering
\begin{tabular}{cccccccc}
& 138 & 139 & 141 & 144 & 146 & 147 & 149 & 151 & 157 & 161 \\
Linked pair 1(a) & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
Linked pair 1(b) & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
Linked pair 2(a) & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
Linked pair 2(b) & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
Unlinked isolate & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
\end{tabular}
\caption{Fragment size in base pairs}
\end{table}

FIGURE 15.2 Fragment data from Figure 15.1 in binary (1/0, present/absent) format, ready for further analysis. For example, a dendrogram is necessary to demonstrate relationships between many isolates.
IV. DISCUSSION

There are several factors that are important in generating clean, reproducible FAFLP profiles. The amount and quality of start DNA is significant, as is performing the two restriction enzyme digestions as separate steps. If the digestions are combined then for the sequenced genomes looked at so far, the in vitro FAFLP results do not match the in silico predictions, whereas following separate restriction digestions they match well. We have also found in our laboratory that if the concentration of the adapters is raised, one or more of the adapter oligonucleotides can act as a PCR primer, producing labeled fragments slightly larger than expected as well as the “true” fragment (e.g., n + 1, n + 2, or n + 3). This corrupts the FAFLP profile. FAFLP is relatively cheap by the standards of other genotyping procedures (approximately $20 per sample), though it cannot match the very modest costs of established phenotypic tests such as phage typing. However, the latter methods mostly have less resolving power than FAFLP. The reagents required for FAFLP are standard in most molecular biology laboratories (i.e., enzymes and oligonucleotides) and FAFLP kits are also available (Applied Biosystems). Unfortunately, the startup costs of any method based on DNA separation on an automated sequencer are high, and none of the platforms currently available are priced low enough for every laboratory to purchase one. The method is, however, relatively fast. Once the method is established, a group of single-species isolates from an outbreak can be compared both with each other and with sporadic unrelated isolates within 48 h.

Fluorescent fragment analysis such as FAFLP requires software capable of sizing fragments and comparing sizes between samples. Fragment analysis software is becoming increasingly available for most sequencing systems, for example from Applied Biosystems (both slab gel and capillary) and Beckman Coulter (capillary). Most fragment software applications require several days’ familiarization but, once trained, operatives find them fairly straightforward. Traditional (i.e., nonfluorescent) methods of analyzing fragment profiles involve scanning the gel image and comparing the 2-D profiles generated using algorithms based on “area under the curve.” FAFLP analysis can also be based on an “area under the curve” measurement, a fast and reasonably accurate method. However, because of its precision, the data can also be analyzed using presence/absence (1/0 scoring or digital scoring) of fragments. This is important for two reasons: (1) the digital data are more easily portable between laboratories and (2) the method appears to be more highly resolving.

FAFLP can equally well be applied to examining the genomic relationships between E. coli and its related enterobacterial species, an area where numerical taxonomy based on combinations of biochemical and other phenotypic tests, and genotyping by sequencing housekeeping genes, are at odds. Future improvements of the FAFLP method will include automation of the whole process. There are several machines already on the market capable of this, starting from extraction of DNA through FAFLP reaction to electrophoretic separation of fluorescent fragments. Other improvements might include “designer” FAFLP, whereby known informative elements can be selectively sampled, or random genome sampling can be performed using genome sequence data to inform the choice of enzymes and FAFLP primers. For example, FAFLP could be performed using enzymes that cut inside a useful marker (such as IS110 in Mycobacterium tuberculosis) to generate FAFLP fragments corresponding to a known element of interest as well as throughout the whole genome. The technique could also be adapted using one of the primers in the PCR directed to a specific element (again such as IS6110) together with FAFLP to generate fragments particular to that element. If total automation can be achieved, FAFLP may be used both for rapid characterization of outbreak isolates and for cumulative accession of isolates of each species of interest with a national database.
Fluorescent Amplified Fragment Length Polymorphism (FAFLP) Genotyping

ACKNOWLEDGMENT

I thank Philip Mortimer for his support and valuable suggestions.

REFERENCES

Global Analysis of DNA Allelic Variations (GADAV) by Specific Enrichment of Mismatches and Selective Amplification of Heterohybrids

Xinghua Pan and Sherman M. Weissman

CONTENTS

I. Introduction.................................................................................................................163
A. Background and Aims.................................................................................................163
B. Outline of the GADAV Strategy.................................................................................164
II. Materials and Methods...........................................................................................164
A. Materials...................................................................................................................164
   1. Immobilized hTDG and mTDG.............................................................................164
   2. DNA Templates and Mm and Pm Oligonucleotide Duplexes...............................165
   3. Oligonucleotides for Adapters and Primers.........................................................165
   4. Other Reagents and Instruments..........................................................................166
B. Experimental Protocol..............................................................................................167
   1. Generation of Heterohybrids and Homohybrids from Two DNA Pools..............167
   2. Enrichment and Separation of Mm Fragments from Pm Fragments.....................168
   3. Selective Amplification of Heterohybrids from a Mixture of DNA Fragments....168
   4. Demonstration and Analysis of the Selected Fragments......................................169
III. Results.....................................................................................................................169
A. Enrichment of an Mm DNA Duplex from Digested pBR322 Pm Fragments.........169
B. Scanning the Polymorphic Fragments in a Human cDNA Pool............................169
IV. Discussion................................................................................................................171
A. Future Developments and Potential Applications..................................................171
B. Summary...................................................................................................................172
Acknowledgments ..........................................................................................................173
References......................................................................................................................173

I. INTRODUCTION

A. BACKGROUND ANDAIMS

Understanding the associations between genetic and phenotypic variation has been a major task of modern molecular biology. As an important part of this process a number of methods have been
developed either for rapid detection of known variation in DNA sequence or for detection of previously unappreciated sequence variation. The latter category of approaches has been difficult to apply to large complex DNA pools on a global scale. E. coli MutHLS proteins have been used to study variations via Identity-by-Descent (IBD) in S. cerevisiae, mouse, and human genomes. However, we and others have not obtained sufficient specificity with these enzymes to detect single nucleotide mutations consistently. Another method, representative differential analysis (RDA), has been successfully used in scanning short or long fragment deletions but it is generally not sensitive to subtle nucleotide variation.

An ideal approach for screening nucleotide variations would fulfill the following criteria: (1) ability to detect all types of single nucleotide variations, (2) high signal to noise ratio, (3) sufficient sensitivity to detect variants present at a low frequency in pooled samples, (4) ability to detect all of the variations in the DNA pools studied in a single experiment, (5) no requirement for prior knowledge of the sequence and its variations, and (6) suitability for automation. Global Analysis of DNA Allelic Variations (GADA) potentially meets all of these requirements. The most significant applications of GADA would be in genetic mapping where it is necessary to distinguish the allelic variations within one DNA pool from that between two different pools. In addition, GADA has many other applications such as screening for heterozygous sites in an individual or for polymorphic sites in a population.

B. OUTLINE OF THE GADA V STRATEGY

Briefly, two pools of DNA are digested with a restriction endonuclease, ligated to a pair of special designed adapters, and then mixed, denatured, and reannealed. The resulting mixture is subject to TDG-mediated enrichment of mismatch fragments (Mm) from perfectly matched fragments (Pm), followed by selective recovery of heterohybrids or homohybrids (Figure 16.1). The output then can be analyzed by gel display or hybridization to genomic or cDNA tiling sequence arrays. This procedure permits global screening of both DNA sequence variants in each pool and of sequence variants that distinguish the sequences in one pool from those in the other.

Our approach uses a special group of mismatch repair enzymes, the thymine (thymidine) DNA glycosylases (TDGs), to separately enrich Mm and Pm DNA duplexes from a complex mixture. Although a number of other DNA glycosylases and mismatch recognizing enzymes for detection of DNA mutations have been described since, these are not suitable for global scanning of DNA pools. The immobilized TDGs used here have two useful properties: first, they selectively remove bases from various internal unpaired nucleotides (when the human and a bacterial TDG are combined they cover all four types of single nucleotide mismatches). Second, they bind to the abasic site thus generated in the presence of EDTA but release the DNA duplexes when magnesium is added. Depending on what adapters were ligated to the original DNA pools, either hetero- or homohybrids can be selectively recovered by PCR. Amplified Mm and Pm fragments can be compared by gel display. Alternatively, array analysis might prove more sensitive and less laborious for complex samples.

II. MATERIALS AND METHODS

A. MATERIALS

1. Immobilized hTDG and mTDG

Human TDG (hTDG) and archaeon Methanobacterium thermoautotrophicum DNA mismatch N-glycosylase (Mig.Mth or mTDG) were produced, fused to glutathione S-transferase (GST), immobilized on Glutathione Sepharose 4B beads (Amersham Pharmacia Biotech), and stored in a buffer containing 50% Glycerol at −80°C. They maintained their activities for up to 2 years.
2. DNA Templates and Mm and Pm Oligonucleotide Duplexes

MspI digested pBR322-DNA (NEB) was 3'-P$^{32}$ labeled with Tag polymerase at 72°C for 5 min in the presence of dGTP and α-P$^{32}$ dCTP.

The mRNA pool was isolated with the oligo(dT)-cellulose type 7 (Amersham) from the Trizol reagent (Gibco BRL) extracted total RNA of the human lymphoblastoid cell line GM12729 (Coriell Cell Repositories). The cDNA pool was synthesized from this mRNA with the SuperScript choice system (Gibco BRL). All steps were accomplished according to the manufacturers’ manual. Two 60 bp duplexes — one perfectly matched duplex (60 Pm) and one containing a G/T mismatched base pair (60 MmCG/GT) — were made. The common bottom oligonucleotide Bo60 was 5'-P$^{32}$ labeled with T4 nucleotide kinase (NEB) before being annealed with different top oligonucleotides to generate the duplexes.$^{14}$

3. Oligonucleotides for Adapters and Primers

A pair of heterohybrid-directing adapters (AdHeA, consisting of the oligonucleotides HeA1 and HeA2, and AdHeB consisting of HeB1 and BeB2) and their corresponding primers [HePa (HeP1), HePb(HeP2), HePaN(HeP1N), HePbN(HeP2N)] were made as shown in Table 16.1. They were designed for selective recovery of the heterohybrids from a mixture of heterohybrids and homohybrids and for further subdivision of the amplicon$^{14}$ based on the PCR Suppression Effect.$^{25,26}$
When the selective nucleotide in the 3' end of the subdivision primer (HePaN or HePbN) was “C,” the third or fourth nucleotide from the 3' end was modified so that it did not match the adapter sequence. Based on our observations and other reports,\textsuperscript{25,27} this increases the specificity of the subdivision.

4. Other Reagents and Instruments

TDG binding buffer:
50 mM Tris-HCl, pH 8.3, 50 mM KCl, 5 mM EDTA, 0.2 mM ZnCl\(_2\), 1 mM DTT, 0.25 mg/ml BSA.

TDG Washing buffers:
The same as TDG binding buffer, but with the separate addition of LiCl in concentrations of 100 mM, 200 mM, 300 mM, 450 mM, and 600 mM.

TDG Elution buffers:
50 mM Tris-HCl pH 7.5, 50 mM KCl, 0.2 mM ZnCl\(_2\), 1 mM DTT, 0.25 mg/ml BSA, and various amounts of MgCl\(_2\) in concentrations of 30 mM, 60 mM, 100 mM, 200 mM, or 300 mM.

FIGURE 16.1 (CONTINUED) (B) Separation and enrichment in bulk of the Mm from the Pm fragments. The mixture generated above was loaded as the input onto the TDG-immobilized beads, bound in the presence of EDTA, washed with increasing concentrations of salt, and eluted with magnesium. This process may be repeated for successive enrichments, in which the previous Pm output is subject to redenaturation and annealing to deplete mismatch fragments. Then, Mm fragment pool and Pm fragment pool are obtained. Finally, heterohybrid pool (AB) and two homohybrid pools (AA and BB) from both Mm and Pm fragment pools can be selectively recovered, and totally six different fragment pools can be obtained. The adapter part of the fragments is not shown here.
**TABLE 16.1**

Sequences for the Heterohybrid-Selective Adapters and Primers

<table>
<thead>
<tr>
<th>Heterohybrids-Selective Adapters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AdHeA</strong></td>
<td></td>
</tr>
<tr>
<td>HeA1: 5′-GTAATACGACTCTATAGGGCTCGAGCGGCCGCCGAGGT</td>
<td></td>
</tr>
<tr>
<td>HeA2: 5′-GATCACCTGCCC</td>
<td></td>
</tr>
<tr>
<td><strong>AdHeB</strong></td>
<td></td>
</tr>
<tr>
<td>HeB1: 5′-TGTAACGTGAAGACGACAGAAAGGGCAGTGGTGAGGGCGGT</td>
<td></td>
</tr>
<tr>
<td>HeB2: 5′-GATCACGGCGCCCTCCG</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Global Primers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HePa: 5′-GTAATACGACTCTATAGGGC</td>
<td></td>
</tr>
<tr>
<td>HePb: 5′-TGTAACGTGAAGACGACAGAA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subsetting Primers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HePaN: 5′-GCGGCCCGCAGGTGATCN</td>
<td></td>
</tr>
<tr>
<td>HePbN: 5′-GCGGCCCGCAGGTGATCN</td>
<td></td>
</tr>
</tbody>
</table>

where the “N” is one nucleotide or a combination of two or three nucleotides.

When “N” in the subsetting primers above is “C”, then,

<table>
<thead>
<tr>
<th>Heterohybrid-Selective Adapters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HePaC</strong></td>
<td></td>
</tr>
<tr>
<td>HePaC: 5′-GCGGCCCGCAGGTGTTCC</td>
<td></td>
</tr>
<tr>
<td>HePbC: 5′-GCGGCCCGCAGGTGTTCC</td>
<td></td>
</tr>
</tbody>
</table>

TDG Perfect-Match Enrichment buffer:

50 mM Tris-HCl pH 7.5, 50 mM KCl, 20 mM EDTA, 0.2 mM ZnCl₂, 1 mM DTT, 0.25 mg/ml BSA.

Titanium Taq DNA polymerase (BD Biosciences Clonetech), T4 DNA ligase and restriction endonucleases (New England Biolab), Wizard DNA clean up system (Promega) or Centricon model 100 (Amicon), GeneAmp PCR system 9600 (Perkin Elmer), sequencing gel system model S2001 (Gibco BRL), Liquid Scintillation Counter (Perkin Elmer), centrifuge, rotator, and isotope protection equipment were used in these experiments.

**B. EXPERIMENTAL PROTOCOL**

1. **Generation of Heterohybrids and Homohybrids from Two DNA Pools**

In a pilot experiment, two cDNA pools “A” and “B,” which were represented by two aliquots of a common human lymphoblastoid cDNA pool, were digested with Sau3AI in the provided buffer at 37°C for 3 h, purified with phenol-chloroform, and precipitated with 1/5 volume 3 M NaAc pH 5.0 and 2.5 volume of ethanol. The resulting DNA restriction fragment pools were separately ligated to the adapters AdHeA and AdHeB with T4 DNA ligase to form the constructs AdHeA-DNA-AdHeA and AdHeB-DNA-AdHeB. The ligation was stopped by adding 20 mM EDTA and heating at 65°C for 20 min. These two pools of constructs were then mixed together, extracted with phenol chloroform, precipitated with ethanol, suspended in 4 μl 3× EE buffer (30 mM EPPS [N-(2-Hydroxyethyl) piperazine-N’-(3-propanesulfonic acid); pH8.0] (Sigma-Aldrich) and 3 mM EDTA) and 1 μl of 5 M NaCl, carefully and thoroughly mixed by pipetting, overlaid with one drop of mineral oil, heated at 94°C for 2 min, quickly cooled to 80°C, then slowly cooled (over 5 h) to 67°C and held at 67°C for 20 h, and finally slowly cooled down to 60°C. This was executed in a thermocycler with a temperature profile of 94°C × 5 min, followed by cooling in one degree steps
for 5 min per step from 80 to 77°C, then 20 min per step from 76 to 71°C, then 60 min per step from 70 to 68°C, 20 h at 67°C, 60 min at 66°C, 65°C, and 64°C, and 20 min at each degree of temperature from 63 to 60°C. The annealing time might be significantly shortened when a much less complex DNA pools such as a PCR amplified subpool was tested. The sample was then diluted to 200 µl, and supplied with PCR components including 1 x buffer, each 100 µM 4 dNTPs, and 1 µl Titanium Taq DNA polymerase (BD Biosciences Clonetech) without primers, and heated at 75°C for 10 min followed by 70°C for 5 min to fill in the DNA ends. The reaction was stopped by adding EDTA 20 mM, followed by purification with the Wizard DNA Clean Up system (Promega) or Centricon Model 100 (Amicon).

2. Enrichment and Separation of Mm Fragments from Pm Fragments

An 80 µl suspension of hTDG or mTDG beads was equilibrated with 500 µl TDG binding buffer and 20 µg to 40 µg HaeIII digested human genomic DNA (other perfectly matched duplex DNA pools would also be acceptable) as a background-blocking carrier was loaded and rotated gently at room temperature for 1 h. About 1 to 10 µg (usually 2 to 5 µg) of the mixture of DNA hybrids made above was loaded into this Eppendorf tube together with a monitoring DNA mixture (1 µg or less, as long as it provided a measurable isotope signal) composed of P32 labeled 60 Pm and 60 MmCG/GT (1:1). The mixture was then rotated gently at room temperature for 6 h or overnight. Two to four cycles of washing were carried out, each with 1000 µl of freshly prepared washing buffer with increasing LiCl (NaCl might also be used, but purification of the output is then more difficult) in concentrations from 100 mM to 600 mM until the P32 isotope signal left in the Eppendorf tube was reduced to 4 to 5% of the total value estimated by Cerenkov radiation, measured in a liquid scintillation counter (Perkin–Elmer). The mixture was then equilibrated with the washing buffer, without EDTA, NaCl, or LiCl, at room temperature for 1 h. Finally 2 to 4 cycles of elution were accomplished by application successively of 1000 µl fresh elution buffer with increasing MgCl2 from 30 to 300 mM. Usually 95% of the remaining P32 signal could be recovered. Each cycle of washing or elution was performed with incubation at room temperature for 4 to 6 h, centrifugation at 1500 rpm for 2 min, and removal of the supernatant. The eluates represented the Mm DNA duplex(es), and could be reloaded to the TDG beads for further enrichment when a higher specificity was desired.

For further enrichment of the Pm duplexes, the initial flow-through portion (supernatant) obtained above before the first washing was phenol-chloroform-ethanol extracted, denatured, and reannealed as described above without filling-in, then mixed with 80 µl of a freshly prepared suspension of TDG beads equilibrated with 500 µl TDG perfect-match enrichment buffer and 20 µg HaeIII digested human genomic DNA, rotated gently at room temperature for 6 h or overnight, and centrifuged at 1000 rpm for 2 min. The supernatant was collected as the Pm duplexes. Fractions from different steps were phenol chloroform extracted, ethanol precipitated, and dissolved in 20 µl of 10 mM Tris-HCl, pH 7.5, 1 mM EDTA for further analysis.

3. Selective Amplification of Heterohybrids from a Mixture of DNA Fragments

The Mm and Pm output DNA samples were adjusted to an appropriate volume (10 to 100 µl) according to the effective DNA template concentration (the background-blocking DNA and the excess adapters were not taken into account), as evaluated by a trial PCR. From each, 2 µl was taken and mixed with PCR components including 1 x buffer, 0.1 µM each of primers HePa and HePb, 0.1 mM each of 4 x dNTPs, and 1/250 volume (0.1 µl for 25 µl total PCR volume) Titanium or Advantage Tag DNA polymerize (BD Biosciences Clonetech), and heated at 75°C for 10 min in a thermocycler. To recover heterohybrids, this reaction was immediately connected to a thermal cycle profile for 14 to 24 cycles, consisting of 94°C for 40 sec, 68°C for 50 sec, and 70°C for 50 sec. On a 1.5% agarose gel, most of
the PCR products would be in the 150–800 bp range. Otherwise, the PCR mixture components and template’s concentration in the mixture should be adjusted. When a single primer (HePa or HePb) is used together with appropriate concentration of key PCR components (especially 0.5 to 1 μm primer, 0.25 mM Δx dNTPs and 0.5 μl Tag polymerase for 25 to 35 cycles of PCR), a given homohybrid (AA or BB) corresponding to the primer used will be produced. To reduce the complexity of a DNA pool, or label and further amplify the amplicons (described below), the PCR product obtained was diluted 10 to 100 fold, and different combinations of primers (HePaN or HePa, and HePbN or HePb) were used in a second round of PCR. In this step, the PCR mixture contained 0.4 μM primers, 0.25 μM each of 4 x dNTPs, and 1/100 volume Tag polymerase, and 20 to 25 thermal cycles were performed.

4. Demonstration and Analysis of the Selected Fragments

Initially the DNA variation scanning and heterohybrid amplification results were demonstrated via gel display. For this purpose, the PCR was accomplished as described above except one primer was 5′-P32 labeled with T4 nucleotide kinase and γ-P32 ATP. Alternatively, the product could be labeled by in-strand incorporation of α-P32 dCTP. The heterohybrids of Mm and Pm were separately amplified. Each 5 μl PCR product was mixed with 5 μl standard sequencing stop solution and heated at 94°C for 3.5 min. A quantity of 3 μl of each product was loaded on a 6 to 8% polyacrylamide/urea gel.

The candidate mismatch fragment bands were recovered by needle puncture and amplified with the same primers used to generate the sample for gel display. The PCR products were digested with Sau3AI and cloned into pBluescript II SK(-) (Stratagene) or TOPO TA Cloning vector (Invitrogen). Inserts of single clones were PCR amplified with the flanking primers of the vectors and transferred to duplicate nylon membranes (Hybond-N+, from Amersham Pharmacia Biotech). The pools of Mm and Pm output cDNA fragments from the column were P32 labeled via PCR and separately hybridized to the membranes. When necessary, the candidate sequences containing an internal mismatch were used to design primers and to confirm the existence of the nucleotide variation in the original DNA pools via PCR-SSCP or direct sequencing.

For high-throughput identification of the Mm and Pm output fragments, hybridization was performed to genomic tiling arrays of PCR fragments (to be described elsewhere).

III. RESULTS

A. ENRICHMENT OF AN MM DNA DUPLEX FROM DIGESTED pBR322 PM FRAGMENTS

Large DNA fragments showed some nonspecific affinity for the beads. To test the effectiveness of the enrichment of Mm fragments from the background of large Pm fragments, the labeled MspI-digested pBR322 (5μg), containing fragments up to 622 bp in length, was mixed with labeled 60 MmCG/GT and applied to the immobilized mTDG beads. After one cycle, the mismatch-enriched eluate showed 10- to 30-fold enrichment for the 60 MmG/T compared to larger plasmid fragments in the input DNA mixture, gel displayed and measured in a PhosphoImager (Molecular Dynamics). Conversely, this Mm fragment was relatively depleted (10- to 30-fold less) from the initial flow-through portion even without further enriching Pm fragments (Figure 16.2). Therefore the ratio of Mm/Pm in the eluate to the ratio of Pm/Mm in the flow-through portion was 100- to 900-fold. The addition of a 25-fold excess unlabeled digested genomic DNA as a carrier improved the specificity.

B. SCANNING THE POLYMORPHIC FRAGMENTS IN A HUMAN cDNA POOL

A human cDNA pool was analyzed for allelic variations in the coding sequences. Five μg cDNA was digested with Sau3AI and heterohybrid-selecting adapters added as described above. After
Mm enrichment with immobilized hTDG, the output pools of Mm and Pm fragments were amplified with a radioactive DNA primer and the resulting heterohybrid fragments fractionated by electrophoresis on an 8% acrylamide sequencing gel (Figure 16.3, panel A1). The complexity of the mixture of fragments was also reduced in test runs by using DNA primers whose 3' end extended beyond the Sau3AI recognition sequence by one or two specifically chosen bases (Figure 16.3, panel A2). Because Mm and Pm DNA pools were amplified to the same final yield of total product, while there were fewer Mm than Pm fragments, the relatively high intensity of bands in the Mm lanes did not reflect the relative depletion or enrichment of these fragments.

Approximately 40 bands that appeared to be enriched in the Mm portion were recovered from the gels by PCR. Because these fragments had symmetric adapters, they could not be sequenced directly but had to be subcloned. However, the subdivided PCR products prepared with different selective nucleotides at the ends of the primers could be sequenced directly. On average, four clones from each band were sequenced. The resulting fragments included DNA sequences from the MHC Class I and Class II alleles, a number of ribosomal protein, mitochondria, ubiquitins, and actins (70%), and from a variety of single copy genes (30%). Several different types of fragments were often recovered from a single band, both because of the complexity of the cDNA mixtures and the impurity of the gel-excised bands. To analyze the enrichment of mismatch fragments, individual cloned fragments were amplified by PCR and subjected to Southern blotting with probes.
prepared from either the total Mm or total Pm match DNA pools (Figure 16.3B). The result showed that a large fraction of the fragments were enriched in the mismatch pool, although the degree of enrichment varied over a fairly wide range.

IV. DISCUSSION

A. FUTURE DEVELOPMENTS AND POTENTIAL APPLICATIONS

We have developed a relatively quick procedure for screening an entire human cDNA pool for single nucleotide variations, with potential for high-throughput. This method can be applied to any DNA pool or DNA pool pair, whether derived from a single individual or a population. In addition, all the identical DNA fragments shared by individuals within a pool potentially can be obtained. The strategy for separating heterohybrids from homohybrids is relatively simple and has satisfactory selectivity. A similar strategy was demonstrated later to be able to amplify the homohybrids under certain conditions. Nevertheless, it might be possible to improve the reliability for heterohybrid selection by
modifying the oligonucleotide adaptors, such as adjusting the length of adapters, designing different matches or mismatches for the primers, or using differentially methylated adapters. In addition, rolling circle amplification\(^2\) may be adopted for highly selective recovery of heterohybrids.

We used a gel display method to demonstrate the feasibility of the method. As an alternative, tiling arrays of unique genomic or cDNA fragments could increase throughput for large-scale applications.

A remaining challenge for GADA V is the formation of DNA duplexes by strands from different members of a family of closely related sequences, such as the MHC Class I genes. This could be overcome by hybridization to genomic or cDNA tiling arrays. This mismatch between loci could also be distinguished from the allelic sequence variation when a self-reannealed DNA is executed as a control if the DNA sample is from haploid genome. When this approach is applied to human or other mammalian genomic DNA, reducing the complexity by a prior isolation of subsets of the DNA would be desirable. This could be achieved by physical separation and by enzymatic subdivision of groups of DNA fragments. Obviously, other restriction enzymes recognizing 4 nucleotide could be used besides Sau3AI to digest the long DNA fragments. Maximum coverage should be obtained by using several enzymes in parallel experiments. To increase the Mn to Pm ratio, endonuclease V and VII and other glycosylases might be useful to combine with TDGs to deplete Mn fragments from the candidate Pm pool. With these enzymes, other subtle DNA variations such as small deletions/insertions and bi-, tri-, and other multiple nucleotide substitutions might also be covered. Biotinylated hydroxy-lamine,\(^2\) an abasic endonuclease, or DNA polymerase beta might also be introduced to remove the random abasic sites generated during DNA manipulation and further improve the signal-to-noise ratio.

This approach can be applied for exhaustive screening of single nucleotide variations for a whole simple or medium complexity DNA pool, a subpool, or a given chromosome region. It is immediately applicable for DNA variation scanning in microorganisms such as bacterial or yeast genomes, where the genomes are haploid and quite simple. Again it might contribute to the discovery of new SNPs (single nucleotide polymorphisms).

In complex genomic DNA scanning, this approach could detect areas of loss of heterozygosity (which would appear as regions without detectable polymorphism) as well as somatic single base mutations arising in malignancies and mutations arising by large-scale mutagenesis in inbred model organisms.

The method suggested here may be applicable to the study of certain human genetic variants, including sporadic dominant mutations that prevent fertility, familial dominant mutations, and recessive mutations. Combined with quantitative microarray hybridization and DNA pooling strategies which can be used for allele-frequency estimation, this approach could be used to screen the nucleotide variations in terms of allele frequency distortion for any phenotype.

### B. SUMMARY

Efficient global scanning of nucleotide variations in DNA sequences between related, complex DNA samples remains a challenge. In the present report we present GADA V as an approach to this problem. We have employed two immobilized thymidine DNA glycosylases which in combination would cover all of the four groups of single nucleotide mismatches to capture and enrich DNA fragments containing internal mismatched base pairs and separate these fragments as a pool from perfectly base-paired fragments as another pool. Enrichments of up to several hundredfold were obtained with one cycle of treatment. We have employed a strategy for selective amplification of heterohybrids which can also be used for selective amplification of homohybrids. By combining these methods together, the single nucleotide variations either between two DNA pools or within one DNA pool can be obtained in one process. This approach has been applied to the total cDNA from a human cell line, and has several potential applications in global scanning of nucleotide variations or polymorphisms in a moderately complex DNA pool, such as microorganismal genomes, and in mapping simple or complex genetic traits in complex genomes such as those of human and mouse.
ACKNOWLEDGMENTS

We thank Dr. R. Lasken (Molecular Staging Inc.) for management throughout this project. This work received financial support from National Cancer Institute Grants R21 CA088326 and R33 CA88326.

REFERENCES


Quantitative DNA-Methylation Analysis by the Bisulfite Conversion Method

Osman El-Maarri, Manfred Kuepper, Johannes Oldenburg, and Jörn Walter

CONTENTS

I. Introduction ................................................................................................................. 175
II. Materials .................................................................................................................... 176
   A. Bisulfite Treatment ................................................................................................. 176
   B. Purification of Bisulfite-Treated PCR Product ........................................................ 176
   C. SNuPE Reaction and HPLC Analysis ..................................................................... 176
III. Bisulfite Treatment and Amplification ....................................................................... 177
   A. Principle of the Bisulfite Conversion Reaction ....................................................... 177
   B. Protocol for Treatment of Intact Cells ................................................................... 177
   C. Protocol for Treatment of Isolated DNA ............................................................... 178
   D. Selective Amplification of Converted DNA ............................................................ 178
      1. Primer Design .................................................................................................... 178
      2. Optimizing PCR Conditions ............................................................................. 179
   E. Troubleshooting ..................................................................................................... 179
   F. Drawbacks of the Method ..................................................................................... 180
IV. Quantification ............................................................................................................. 180
   A. Cloning and Sequencing ....................................................................................... 181
   B. Analysis by SIRPH (SNuPE-IP RP HPLC) ......................................................... 181
      1. Protocol Outline ................................................................................................ 181
      2. Guidelines for Designing SNuPE Primers ......................................................... 183
      3. Analysis of the Results ..................................................................................... 183
      4. Troubleshooting ............................................................................................... 183
      5. Advantages and Drawbacks of the SIRPH Method ........................................... 184
V. Conclusions................................................................................................................ 184

Acknowledgments ........................................................................................................ 184
References ...................................................................................................................... 184

I. INTRODUCTION

The most advanced and detailed method to analyze cytosine methylation in genomic DNA is the bisulfite-based sequencing introduced by Frommer and colleagues in 1992. It is based on bisulfite treatment which converts unmethylated cytosine to uracil, while methylated cytosines remain unchanged under the experimental conditions used. The great potential of this method was quickly
recognized, and a number of improvements of the original protocol were published, mostly to achieve better sensitivity and conversion rates (a systematic analysis of reaction parameters is given in Jost et al.). Other improvements focused on use of the method on imbedded tissues and intact cells. In 1996, we published a protocol in which the whole reaction is performed on material — DNA or cells — embedded in low-melting-point (LMP) agarose. This approach proved to be superior to other protocols in several aspects. It facilitates sample handling by eliminating precipitation steps and does not require DNA isolation out of cells or early embryos. It thereby reduces loss of material and allows the assessment of minute sample quantities. Furthermore, the physical trapping of the DNA within the agarose matrix prevents renaturation and thus helps to keep the DNA single-stranded, an indispensable prerequisite for an efficient conversion reaction. After generating a PCR product, the next step in a bisulfite analysis is to quantitate the ratio of cytosine to thymine content at a specific CpG site. To address this, many methods have been developed. In this chapter, we will describe in detail the traditional cloning and sequencing method (Section IV.A) and compare this with our recently published SIRPH protocol (Section IV.B).

II. MATERIALS

A. Bisulfite Treatment

The following reagents are required for this step:

- Phosphate-buffered saline (PBS)
- Trypsin (Biochrom), 0.25% (w/v) in PBS
- Mineral oil (heavy white mineral oil, Sigma)
- LMP agarose (SeaPlaque agarose, FMC), 2% in PBS (Section III.B) or water (Section II.C)
- Tris-EDTA (TE) buffers pH 7.0 and pH 8.0: 10 mM Tris-HCl, 1 mM ethylene-diaminetetraacetic acid (EDTA)
- Phenylmethylsulfonyl fluoride (PMSF, Sigma) 40 µg/ml in TE buffer pH 7.0
- 2.5 M sodium bisulfite solution pH 5: dissolve 1.9 g of sodium metabisulfite (Merck) in a mix of 2.5 ml H2O and 750 µl of 2 M NaOH; dissolve 55 mg of hydroquinone (Sigma) in 500 µl H2O at 50°C; and mix the two solutions. Sodium bisulfite and NaOH solutions should always be prepared freshly. Bisulfite and hydroquinone are light sensitive and must be protected from light throughout the whole procedure.
- Lysis solution: 10 mM Tris-HCl, 10 mM EDTA, 1% sodium dodecyl sulfate (SDS), 20 µg/ml proteinase K (Roche).
- Restriction enzymes and buffers.
- NaOH 0.1 M, 0.4 M, 2.0 M

B. Purification of Bisulfite-Treated PCR Product

PCR products can be purified either by gel extraction (e.g., QIAquick Gel Extraction Kit from Qiagen, Cat. No. 28704) or by treatment with a mixture of exonuclease I and shrimp alkaline phosphatase (ExoSap-IT from Amersham, US78201).

C. SNuPE Reaction and HPLC Analysis

Thermosequenase, dTTP and ddCTP are from Amersham (catalogue numbers E79000Y, 27-2081-01, and 27-2061-01, respectively).

For dHPLC analysis, we are using the Wave system from Transgenomics, together with the IP RP HPLC column, the DNA-Sep, and the TEAA buffer (Part no. 450181, Part no. 553303). Acetonitril is from ROTH (Art 8825.2), dHPLC grade water from Merck (Cat no. 1.15333.2500).
III. BISULFITE TREATMENT AND AMPLIFICATION

A. PRINCIPLE OF THE BISULFITE CONVERSION REACTION

The main principle of this method is the selective conversion of cytosine residues to uracil in the presence of sodium bisulfite. This reaction requires three major steps (Figure 17.1): (1) sulfonation, (2) deamination, and (3) desulfonation.\(^5\)

1. Reversible sulfonation of cytosine residues to cytosine-6-sulfonate. This reaction is favored at low pH and low temperature; at 0°C the equilibrium state is reached within 20 min.
2. Irreversible hydrolytic deamination of cytosine-6-sulfonate to uracil-6-sulfonate. This reaction is favored at higher concentrations of sodium bisulfite and at higher temperatures; the pH optimum is between pH 5 and 6.
3. Reversible desulfonation of uracil-6-sulfonate to uracil. The elimination reaction is favored at higher pH.

Since cytosines in double-stranded, nondenatured DNA are almost inert to the reaction, it is crucial to maintain the DNA single-stranded. Under the described conditions, the reaction is highly selective and almost complete for nonmethylated cytosines, whereas 5-methylcytosines remain nearly entirely unconverted.\(^6\)

B. PROTOCOL FOR TREATMENT OF INTACT CELLS

The possibility to investigate intact cells without prior DNA extraction allows the assessment of limited amounts of tissue or even few cells. However, when larger samples are available we recommend isolating the genomic DNA and following the protocol for treatment of isolated DNA (see Section III.C).

Detailed protocol:

1. Tissue samples should preferably be trypsinized to obtain a single-cell suspension. Omit this step if individually collected cells, such as oocytes and zygotes, are being analyzed.
2. Wash and recover cells in PBS at a maximum density of 60 cells/µl.
3. In a 2-ml Eppendorf tube, mix 3 µl of the cell suspension with 6 µl of hot (80°C) 2% (w/v) LMP agarose, prepared in PBS.
4. Add 500 µl of heavy mineral oil, incubate in a boiling water bath for 20 min, and transfer to ice for an additional 30 min to solidify the agarose/cell mixture.
5. Incubate the agarose bead in 500 µl of the lysis solution, overlaid with the mineral oil, at 37°C overnight.
6. Remove the lysis solution and the oil and inactivate proteinase K by adding 500 µl of 1x TE pH 7.0 containing 40 µg/ml PMSF. Incubate 30 min at room temperature. This step is optional.

![Figure 17.1](www.taq.ir)

**FIGURE 17.1** Principle of the bisulfite reaction. (I) Sulfonation of the cytosine at position C6. (II) Irreversible hydrolytic deamination at position C4, generating 5-sulfonate-uracil. (III) Desulfonation under alkaline conditions. Methylation at position C5 impedes sulfonation at position C6 (step I).
7. Remove all solution and wash with TE (pH 8.0) for 2 × 15 min (two separate incubations insure the complete deactivation of PMSF by TE pH 8).
8. Calibrate against 100 µl of the restriction buffer for 15 min.
9. Remove the solution and add 100 µl of 1x restriction buffer containing 20 units of restriction endonuclease (selected to cut outside the amplified region) and incubate overnight at 37°C. As an alternative, add 50 U and digest for 1 h at 37°C.
10. Remove the restriction buffer and incubate with 500 µl of 0.4 M NaOH for 2 × 15 min.
11. Wash with 1 ml of 0.1 M NaOH for 5 min.
12. Remove all solution and add 100 µl of 1x restriction buffer containing 20 units of restriction endonuclease (selected to cut outside the amplified region) and incubate overnight at 37°C. As an alternative, add 50 U and digest for 1 h at 37°C.
13. Boil for 5 to 10 min in a water bath to inactivate the enzyme and to thoroughly denature the fragmented DNA.
14. Chill quickly on ice and spin down briefly to collect any condensate.
15. Add 4 µl of 0.1 M NaOH (final concentration 0.3 M NaOH) and incubate for 15 min at 50°C.
16. Mix with 2 volumes (50 µl) of hot (liquid) 2% (w/v) LMP agarose (prepared in water).
17. Pipette 1 ml of the 2.5 M sodium bisulfite solution into a 2-ml Eppendorf tube and overlay with 750 µl of heavy mineral oil. It is recommended to keep tubes on ice for 30 min to enhance bead-formation in the prechilled oil.
18. Pipet up to seven 10 µl drops of the DNA-agarose mixture into ice-cold mineral oil to form beads. Remember that no bead should contain more than 100 ng of DNA. Make sure that all beads entered the aqueous (lower) phase; if necessary, they can be pushed into the bisulfite solution using a pipette tip.
19. Leave on ice for 30 min.
20. Incubate at 50°C for 3.5 h.
21. Remove all solution; wash 1 ml TE pH 8.0 for 2 × 15 min.
22. Add 500 µl of 0.2 M NaOH and incubate for 2 × 15 min.
23. Remove NaOH solution and wash with 1 ml TE pH 8.0, 3 × 10 min. Store in few µl of TE buffer at 4°C. The beads are stable for at least several months.
24. Prior to PCR amplification, wash the beads with pure water for 2 × 15 min.

C. Protocol for Treatment of Isolated DNA

In cases where the material to be analyzed is abundant, it is recommended to isolate chromosomal DNA, according to Miller et al.7, and proceed with the following protocol:

1. Digest the genomic DNA with a suitable restriction enzyme (which does not cut within the region to be amplified) in a volume of 21 µl. In order to achieve a complete bisulfite conversion, the amount of DNA should not exceed 700 ng, corresponding to no more than 100 ng contained in each agarose bead formed later (see step 8).
2. Boil for 5 to 10 min in a water bath to inactivate the enzyme and to thoroughly denature the fragmented DNA.
3. Chill quickly on ice and spin down briefly to collect any condensate.
4. Add 4 µl of 2 M NaOH (final concentration 0.3 M NaOH) and incubate for 15 min at 50°C.
5. Mix with 2 volumes (50 µl) of hot (liquid) 2% (w/v) LMP agarose (prepared in water).
6. Pipette 1 ml of the 2.5 M sodium bisulfite solution into a 2-ml Eppendorf tube and overlay with 750 µl of heavy mineral oil. It is recommended to keep tubes on ice for 30 min to enhance bead-formation in the prechilled oil.
7. Pipet up to seven 10 µl drops of the DNA-agarose mixture into ice-cold mineral oil to form beads. Remember that no bead should contain more than 100 ng of DNA. Make sure that all beads entered the aqueous (lower) phase; if necessary, they can be pushed into the bisulfite solution using a pipette tip.
8. Leave on ice for 30 min.
9. Incubate at 50°C for 3.5 h.
10. Remove all solution; wash 1 ml TE pH 8.0 for 2 × 15 min.
11. Add 500 µl of 0.2 M NaOH and incubate for 2 × 15 min.
12. Remove NaOH solution and wash with 1 ml TE pH 8.0, 3 × 10 min. Store in few µl of TE buffer at 4°C. The beads are stable for at least several months.
13. Prior to PCR amplification, wash the beads with pure water for 2 × 15 min.

D. Selective Amplification of Converted DNA

1. Primer Design

The following guidelines concern primer design and conditions for the selective and efficient amplification of only the bisulfite modified DNA.
1. The method described here can be used for studying genomic imprinting through differential methylation of the parental genes. In this case, the region to be amplified should contain sequence polymorphisms that allow the identification of the parental alleles after bisulfite treatment (e.g., G to A transitions). This may limit the analysis to either the lower or upper strand.

2. It is helpful to generate a “bisulfite-converted” DNA sequence by computational substitution of all C residues for T residues except at CpG sites. Such data file can be used by any primer-designing software to test the selected primers to avoid hairpin structures, false priming sites, and possible primer dimers. Alternatively MethPrimer (http://itsa.ucsf.edu/~urolab/methprimer), a freely available program developed by M.-C. Li and R. Dahiya, will perform this task automatically.

3. Overlapping of the primers with CpG dinucleotides should be strictly avoided, especially at the 3’ end of the oligos. This is important to ensure that the PCR amplification is independent from the conversion efficiency or the methylation status of these CpG sites that overlap with the primers.

4. For a robust yet specific primer annealing a primer length of 25 to 30 nucleotides is advised.

5. To ensure selective amplification of only converted DNA primers should be located, if possible, in an originally cytosine-rich region (but not CpG).

6. Extensive T and A stretches, which are common to bisulfite-treated DNA, should be (if possible) avoided in both primers to minimize the formation of primer dimers.

2. Optimizing PCR Conditions

The bisulfite treatment changes all non-methylated cytosine residues into uracils and thus reduces the sequence specificity and selectivity for primer annealing. This requires careful and thorough optimization of the PCR conditions.

1. The recommended length of the products should not exceed 600 to 700 bp. Longer fragments may be difficult to amplify due to depurination of the DNA at low pH and the resulting fragmentation.

2. The specificity and sensitivity can be increased by applying a nested, or at least semi-nested, PCR approach. However, when using such an approach several products from independent experiments should be analyzed to insure the reproducibility of the results.

E. Troubleshooting

1. During formation of the beads, the agarose should be kept at 50 to 65°C. If it becomes too cold, some or all of the DNA-agarose mixture may solidify inside the pipette tip. Also, the pipette tip should be inserted into the cold mineral oil only slightly with the very tip and the mixture should be discharged rapidly.

2. To avoid collision and hence fusion of beads during formation, they should be pipetted into opposite sides of the tube and the number of beads should not exceed four per tube.

3. If the agarose beads should dissolve after entering the bisulfite solution, the DNA-agarose mixture has not been mixed homogeneously or the layer of mineral oil is not cold enough. To prevent that, the tubes containing mineral oil should be preincubated on ice for at least 20 min. Moreover, only heavy mineral oil of pure quality (e.g., from Sigma) should be used. If the problem persists, we recommend increasing the concentration of the LMP agarose. Alternatively, beads can be formed on a clean sheet of parafilm kept on ice and transferred manually into the bisulfite solution.
4. Ice crystals appearing during incubation of the bisulfite-hydroquinone solution on ice will not affect the results.

5. If unconverted sequences are observed frequently, the following should be considered:
   (1) Primers are not selective enough for converted DNA. The primers should be located in a C-rich region to increase the selectivity of amplification toward fully converted sequences. (2) An excess of DNA in the reaction may trigger an incomplete conversion. We recommend not to exceed 100 ng per bead. (3) The DNA was not properly denatured. Make sure that denaturation steps and desulfonation steps are carried out using freshly prepared NaOH solution and sodium bisulfite solution.

6. Failure of PCR amplification may be caused by (1) inefficient bisulfite conversion; (2) insufficient amount of template DNA; (3) size of the desired product — try to amplify a smaller fragment; (4) low sensitivity of the amplification. In the latter case, a nested PCR approach is recommended or, alternatively, the use of a different set of primers.

7. A systematic analysis by Warnecke et al. nicely demonstrated that for some templates, specially for CpG rich ones, a bias in the PCR reaction might occur such that either a low or highly methylated template DNA is predominantly amplified.

The problem of biased amplification has to be tested individually, and several control experiments should be carried out. First, different templates with a known content of methylated cytosine residues should be mixed in different ratios and the bisulfite treatment and amplification steps carried out as usual. The distribution of nonconverted and converted cytosine residues in the analyzed products will then allow determination of whether, and to what magnitude, a bias has occurred. A strategy to avoid such problems is to perform independent experiments (including different techniques) to analyze the methylation profile of a given template, for example, by the SIRPH technique presented below.

F. DRAWBACKS OF THE METHOD

1. Detailed sequence information about the genomic region of interest is required in order to perform a bisulfite-based methylation analysis.

2. After bisulfite treatment, the former upper and lower strands are not complementary anymore and hence must be analyzed separately. Therefore it is impossible (except for single-cell analysis) to obtain data about the original double-stranded DNA.

3. Amplifications and sequence reactions from the upper and lower strands may not work equally well in all cases.

4. In genomes with non-CpG methylation or asymmetrical methylation patterns, such as those of plants and fungi, it can be difficult to design primers for the PCR amplification of the bisulfite-treated DNA. In this case, primers can be designed that contain either C or T at the respective positions. However, the use of such primers with “wobble” positions greatly reduces the specificity of the PCR reaction and may cause a bias against the amplification of certain not fully converted products.

IV. QUANTIFICATION

The main goal in DNA-methylation analysis is to rapidly and reliably quantify the amount of methylated cytosines (chromosomes) in a given genomic region. The traditional method of cloning and sequencing of bisulphite PCR-products provides the most detailed information. This approach, however, is very time consuming, laborious, and expensive. For many applications the analysis of methylation at selected CpGs will be sufficient. Since there is a great demand for cheap, rapid, and quantifiable methods to perform high-throughput screening of DNA methylation for diagnostic purposes, several alternative methods have been developed in recent years: One of such methods is COBRA, which is based on restriction enzyme digestion of bisulphite PCR products. The use of
this method however is limited since it only allows the analysis of CpG methylation within (newly generated) restriction sites of the bisulphite PCR products. Recently a more flexible method, based on differential hybridization of bisulphite PCR fragments using oligonucleotide-containing chips, was introduced.\textsuperscript{11} While this method allows a high-throughput screening, it requires a high technological laboratory standard and a sophisticated and laborious Chip-design and analysis tools. More flexible methods are based on single nucleotide primer extension SNuPE techniques: Methylight, a real-time PCR-based SNuPE method, is quantitative and highly sensitive but requires special fluorescently labeled primers.\textsuperscript{12} Ms-SNuPE uses incorporation of radioactive nucleotides quantified by autoradiography and requires separation of extension products on acrylamide gels.\textsuperscript{13} Recently, we have developed a cheap, nonradioactive variation of such a SNuPE protocol using ion pair reverse phase high performance liquid chromatography (SIRPH) for methylation detection and quantification.\textsuperscript{4}

A. \textbf{Cloning and Sequencing}

The Bisulfite derived PCR product can be cloned in a TA cloning vector to analyze the methylation patterns of individual clones.

The PCR product has to be amplified with an enzyme that introduces a 3’ flanking dATP to the products, e.g., Taq polymerase. The efficiency with which \textit{Taq} polymerase adds a flanking dATP increases if the last base is either A or T, which can be added to 5’ end of the primer sequence for this purpose. Moreover, primers should be selected to amplify a fragment of 500 to 700 bp for optimal cloning efficiency. After removing excess primers and dNTPs (by gel extraction or column purification; e.g., QIAquick Gel Extraction Kit from Qiagen), the purified product can be used for cloning.

B. \textbf{Analysis by SIRPH (SNuPE-IP RP HPLC)}

Bisulphite PCR products of the region of interest are obtained using the described protocol (see Section III above). The PCR products are then purified (see following protocol) to remove residual PCR oligos and dNTPs. Subsequently, primers immediately 5’ to a CpG site are hybridized to the denatured single stranded PCR product. Temperature cycling using Thermo Sequenase TM in the presence of both ddCTP and ddTTP extends the annealed primers. The ddTTP ( unmethylated CpG) or ddCTP (methylated CpG, see Figure 17.2) extended products are then directly loaded on a dHPLC column (WAVE DNA Fragment Analysis System, Transgenomics). Due to the incorporation of the more hydrophobic ddTTP the retention time of such an extended product is longer compared to products containing ddCTP (Figure 17.1). The amount of the ddTTP and ddCTP extended products can then be quantified by either integrating the area or measuring the height of the peaks and calculation of their percentage ratio using the WaveMaker program (Version 4.1, Transgenomics, see also Section IV.B 3).

1. **Protocol Outline**

The SIRPH protocol for the quantitative methylation analysis at specific CpG sites can be divided into three parts. The first part involving the generation of the bisulphite PCR products (see Section III above), the second is the SNuPE reaction while the third is the separation of the product and analysis on the HPLC. In the following sections we will address in detail the last two parts.

a. **Purify the PCR Product by One of the Two Methods:**

1. Run the product on 1% agarose gel until optimum separation, excise the specific band, and recover the product by using a standard PCR-gel extraction kit. In our experience the QIAquick Gel Extraction Kit from Qiagen (Cat. No. 28704) yields highly pure products with high rates of recovery.

2. Add 2 µl of exonuclease I and shrimp alkaline phosphatase (ExoSap-IT from Amersham (US78201)) to 5 µl of PCR product, and heat at 37°C for 15 min, followed by 15 min, at 80°C to deactivate the enzyme mixture.
Enzymatic treatment is more expensive, but has the advantage that it is rapid and easier to perform, especially when analyzing large number of samples. Gel extraction, on the other hand, is more laborious but it has the advantages of concentrating faint PCR product in smaller volume. It offers also the possibility of isolating the specific product in cases where there are unspecific products.

**b. Set up the SNuPE Reaction in a Total Volume of 20 μl with the Following Components:**

- **Reaction buffer:** 2 μl (10 X buffer)
- **SNUPE Oligos (n):** 1 μl (12.5 pmole solution/for each oligo)
- **Template DNA:** 1 to 5 μl (50 to 100 ng)
- **ddCTP:** 1 μl (1 mM solution)
- **ddTTP:** 1 μl (1 mM solution)
- **Thermo Sequenase n μl** (diluted to 1 unit/μl, n = the number of oligos used for multiplex in the reaction)
- **H₂O:** up to 20 μl
Quantitative DNA-Methylation Analysis by the Bisulfite Conversion Method


c. Amplify Using the Following Program:

Step 1: 94°C for 2 min
Step 2: 92°C for 10 sec
Step 3: 30°C for 30 sec
Step 4: 60°C for 1 min
Repeat steps 2 to 4 for 50 times

d. Run the Products on HPLC

Load 10 to 15 μl of the above product directly on the HPLC machine (Wave from Transgenomics). Set the oven temperature to 50°C and the elution gradient (mixture of buffers A and B) at 0.9 ml/min over 10 min.

When setting up a new assay, run with a wide gradient of 10% (b1) to 60% (b2) buffer B. Most short oligos of 10 to 20 bp should be eluted by this gradient. At a later stage, and dependent on the retention time of the oligos, the gradient can be narrowed down from either left side, right side, or both sides simultaneously to give the best spatial resolution between the oligos.

2. Guidelines for Designing SNuPE Primers

The most important step in the SNuPE-IP RP HPLC is to select and compose suitable sets of oligos. Some recommendations for designing oligos are listed below:

1. The 3’ end of the SNuPE oligo has to be just 5’ (flanking) of the specific CpG site to be studied.
2. Avoid placing the oligo on a T-rich region, as this could increase mispriming and lead to inaccuracy in the methylation measurements.
3. Oligos should not include a CpG site, as this will bias the measurements.
4. Oligos that are too short have higher chance of mispriming. In our experience, oligos as short as 10 bases can still provide accurate data. However, for routine uses we prefer oligos 15 to 18 bases long.
5. For multiplex SNuPE reactions that are run simultaneously on the HPLC, the retention time of the individual oligos and their elongation products should be different. If, for practical reasons, two oligos have to be designed that give similar retention times on the HPLC, we recommend to extend one of the oligos by adding thymidins to its 5’ end. In our experience this addition has neither any effect on the annealing to the template or on the SNuPE reaction. The number of Ts to be added has to be determined empirically, however, each additional T has a stepwise additional retardation effect in a linear fashion.

3. Analysis of the Results

The percent of methylated portion of the DNA can be calculated according to the formula: \( M = \left( \frac{AC}{AC + AT} \right) * 100 \), were AC and AT are the peak areas or peak height of the ddCTP and ddTTP extended oligos, respectively. The WaveMaker software automatically calculates the AC and AT.

4. Troubleshooting

Some problems may be encountered while trying to set up the SIRPH protocol for the first time. Listed below are the most common ones, together with tips to overcome them.
1. The peaks of the oligos and/or their extension are overlapping on the HPLC histogram. Resolution can be improved by adding additional Ts to increase the retention time of certain oligos.

2. Where a slight overlap between two corresponding peaks cannot be avoided more accurate measurement can be obtained from comparing peak heights rather than peak areas.

3. When reloading the same extension product repeatedly, peak retention values should not differ by more than 1 to 3% in a sample that has a theoretical 40 to 60% methylation. However, this error can increase for peaks that are relatively small. In such cases, a higher volume of the SNuPE product should be loaded to minimize the differences between separate runs.

4. For the best reproducible results, it is recommended to leave an interval of about 2 min between the SNuPE peaks and both the injection and washing peak.

5. Advantages and Drawbacks of the SIRPH Method

The SIRPH methylation analysis has the advantage of being quantitative, rapid, nonradioactive and suitable for automation. However, the method has also some drawbacks:

1. In contrast to the cloning and sequencing approach, combinatorial variation in methylation patterns (haplotypes) can not be distinguished.

2. The imprinted patterns of methylation, with about half of the PCR product completely methylated and the other half completely unmethylated, cannot be revealed directly but have to be deduced from the level of methylation at only few sites.

3. In a CpG island where CpG sites are very close to each others (i.e., between 0 and 10 bases), some CpCs cannot be analyzed.

V. CONCLUSIONS

In this chapter, we have described detailed protocols for bisulphite treatment of both extracted DNA and intact cells. As a general strategy, when analyzing a new region for its methylation content, we recommend to do cloning/sequencing for at least one sample. Subsequently, if large number of samples needs to be analyzed, we recommend establishing a high-throughput approach like the SIRPH protocol. The latter protocol is highly reproducible, easy to perform, semiautomated, and has the advantage that large number of samples can be processed in a short time.

ACKNOWLEDGMENTS

We thank Joachim Fischer and Sven Langner from Transgenomic for valuable discussions and technical support, and Ursula Herbiniaux and Natascha Rinker for technical support.

REFERENCES


18 Pyrosequencing Technology

Pål Nyrén, Tommy Nordström, and Baback Gharizadeh

CONTENTS

I. Introduction.................................................................................................................187
II. Materials and Methods.............................................................................................188
   A. PCR Conditions......................................................................................................188
   B. Preparation of DNA Template..............................................................................189
      1. Single-Stranded Template.............................................................................189
      2. Double-Stranded Template...........................................................................189
   C. Pyrosequencing......................................................................................................190
III. Results....................................................................................................................190
   A. Choice of PCR Primers........................................................................................190
   B. DNA Preparation...................................................................................................191
      1. Single-Stranded Template.............................................................................191
      2. Double-Stranded Template...........................................................................191
   C. Pyrosequencing......................................................................................................191
   D. Applications..........................................................................................................192
IV. Discussion................................................................................................................193
Acknowledgments ...........................................................................................................193
References......................................................................................................................194

I. INTRODUCTION

The ability to determine nucleic acid sequences is of utmost importance for the detailed study of biological systems. Pyrosequencing™ is the first alternative to the traditional Sanger method for de novo DNA sequencing.1,2 The Pyrosequencing technology is especially suited for short-read sequencing,3 although the method has recently been improved to allow sequencing of up to hundred bases.4 It is broadly applicable for analysis of single nucleotide polymorphism (SNPs)5–8 and for identification of short DNA sequences used in bacterial,9 fungal (unpublished data), and viral typing.10 In addition, the methodological performance of the Pyrosequencing technology has been used to determine difficult secondary structures,11 mutations,12,13 and for tag sequencing of a selected cDNA library.14

Pyrosequencing employs coupled enzymatic reactions in order to monitor the inorganic pyrophosphate (PPI) released during nucleotide incorporation (Figure 18.1). A sequencing primer is hybridized to a single-stranded DNA template and mixed with the enzymes DNA polymerase, ATP sulfurylase, luciferase, and apyrase, as well as the substrates adenosine 5’ phosphosulfate (APS) and luciferin. Then the first of four deoxynucleotide triphosphates (dNTP) is added to the reaction. DNA polymerase catalyzes the incorporation of the deoxynucleotide into the DNA-strand only if it is complementary to the base in the template strand. Each incorporation event is accompanied by the release of PPI in a quantity equimolar to the amount of incorporated nucleotide. ATP sulfurylase quantitatively converts PPI to ATP in the presence of APS. The produced ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light in amounts
that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction
is detected by a charge-coupled device (CCD) camera and seen as a peak in a Pyrogram™. Each
light signal is proportional to the number of nucleotides incorporated. Apyrase, a nucleotide
dergrading enzyme, continuously degrades ATP and unincorporated dNTPs. When degradation is
complete, another dNTP is added. Addition of dNTPs is performed one at a time. As the process
continues, the complementary DNA strand is built up and the nucleotide sequence is determined
from the signal peaks in the Pyrogram. Pyrosequencing has the advantages of accuracy, flexibility,
simple automation, and parallel processing.

The Pyrosequencing technology is commercially available from the company Pyrosequencing
AB, Uppsala, Sweden (www.pyrosequencing.com). The company’s products include the bench-top
PSQ™96MA System and a high-throughput PTP™ System, which utilize proprietary software and
reagent kits.

II. MATERIALS AND METHODS

A. PCR CONDITIONS

The oligonucleotides Biotin-pUC-5G-UP (5'-GGCTCGTATGTTGTGTAAGA), pUC-5G-DOWN
(5'-CAGCTTCGCTTTGATTACG), pUC-5G-Seq (5'-TTGCTATTACTGAGCTGG), pUC-5T-
UP (5'-GCATGTGTCAGAGGTTTTCA), Biotin-pUC-5T-Down (5'-CATTTATCAGGGTTTAGT-
TGTC), pUC-5T-Seq (5'-CGAAAGCCCGCTCGTACG), pUC-5T-UP3P (5'-GCATGTGTCAG-
GTTTTCA-Phosphate) and pUC-5T-Down3P (5'-CATTTATCAGGGTTTAGTC-Phosphate), were designed for pUC19 vector (New England BioLabs, Beverly, MA). All oligonucleotides
were synthesized and HPLC purified by Interactiva (Ulm, Germany).

PCR reactions were performed using primer pair Biotin-pUC-5G-UP and pUC-5G-DOWN for
amplification of a 241-bp fragment, and pUC-5T-UP and Biotin-pUC-5T-DOWN for amplification
of a 215-bp fragment. For primer pair Biotin-pUC-5G-UP and pUC-5G-DOWN, PCR was carried out in
total volume of 50 µl containing 10 ng pUC19 plasmid; 10 mM Tris-HCl (pH 8.3); 2 mM magnesium
chloride; 50 mM KCl; 0.1% (v/v) Tween 20; 0.2 mM dNTPs; 1 unit of AmpliTaq DNA polymerase
(Perkin–Elmer, Norwalk, CT); and 10 pmol of each primer. A 2-min denaturation step at 94°C was
followed by 32 cycles of amplification with a PTC-200 PCR system (MJ Research, Inc., Waltham,
Pyrosequencing Technology

MA). Each cycle comprised of a denaturation step at 94°C for 40 sec, a primer-annealing step at 56°C for 45 sec, and a chain elongation step at 72°C for 45 sec. The final elongation step was prolonged by 10 min to ensure a complete extension of the amplified DNA. For primer pair pUC-5T-UP and Biotin-pUC-5T-DOWN, PCR was performed in total volume of 50 µl containing 10 ng pUC19 plasmid; 1 unit Taq DNA polymerase (KEBO Lab, Spånga, Sweden); 4 mM magnesium chloride; 0.2 mM dNTPs (Amersham Biosciences); and 10 pmol of each primer. The PCR was carried out in a PTC-200 PCR system. Thermal cycling was performed with an initial denaturation for 10 min at 95°C, succeeded by 50 cycles of denaturation for 45 sec at 95°C, primer annealing for 45 sec at 49°C, and synthesis for 60 sec at 72°C. A final primer extension was conducted for 5 min at 72°C.

B. PREPARATION OF DNA TEMPLATE

1. Single-Stranded Template

   1. Take 50 µl biotinylated PCR product, immobilize onto streptavidin-coated super paramagnetic beads (Dynabeads M-280-streptavidin, Dynal AS, Oslo, Norway) by incubation at 43°C for 30 min.
   2. Place on magnetic stand and remove supernatant. Obtain ssDNA by denaturing the immobilized PCR product in 5 µl 0.1 M NaOH for 5 min. The immobilized strand was resolved in 45 µl annealing buffer (20 mM Tris-acetate (pH 7.6), 5 mM magnesium acetate).
   3. Hybridize ssDNA corresponding to 25 µl PCR product to 10 pmol sequencing primer (pUC-5G-Seq) at 70°C for 3 min followed by incubation at room temperature for 5 min. If desired, the sample can be washed with 50 µl of annealing buffer to remove the unbound sequence primers.

   In the described example, the sequencing was performed in the presence of 2.5 µg E.coli single-stranded DNA binding protein (SSB) (Amersham Biosciences, Uppsala, Sweden).

2. Double-Stranded Template

   For direct Pyrosequencing on dsDNA, it is nececssary to first destroy any primers and dNTPs which might be left over from the PCR reaction and which could function as substrates.

   a. Two-Step Method

      1. Take 10 µl PCR-amplified product, add 2.5 µl of a solution containing 50 mU apyrase (Sigma Chemical Co., St. Louis, MO), 1 mU HPLC purified yeast inorganic pyrophosphatase (PPase) (Sigma Chemical Co.), 2 U E. coli exonuclease I (Amersham Biosciences), 20 mM Tris-acetate (pH 7.6), 2 mM magnesium acetate, 2% BSA, 1 mM DTT. Mix and incubate at 35°C for 20 min.
      2. Inactivate enzymes at 98°C for 3 min.
      3. Add 1 µl (5 pmol) of sequencing primer and anneal at 98°C for 2 min. Transfer the sample rapidly to an ice-water bath.

   For details of this method, see Reference 15.

   b. One-Step Method

      1. Add: 10 µl of the PCR-amplified product
         35 µl of a solution containing 20 mM Tris-acetate (pH 7.6), 2 mM magnesium acetate, 2% BSA, 1 mM DTT
         100 mU apyrase
         10 mU HPLC purified yeast PPase
         20 pmol sequencing primer (pUC-5T-Seq)
20 pmol of blocking oligonucleotide (pUC-5T-UP3P)
20 pmol of blocking oligonucleotide (pUC-5T-Down3P)
Mix and incubate at 35°C for 6 min.
2. Heat reaction mix at 100°C for 1 min and then place immediately in an ice-water bath, or on cold (−20°C) glycerol for 1 min. In the described example, the complete temperature program was performed in a PCR machine (MJ Research PCR system).

For details of this method, see Reference 16.

C. PYROSEQUENCING

Pyrosequencing was performed at 28°C in a volume of 50 µl on an automated bench-top PSQ™96 System. Briefly, reagents are loaded into the cartridge that is inserted into the dispenser. The 96-well plate, containing the DNA samples (ssDNA or dsDNA prepared as described above) annealed to sequencing primers (final volume 45 µl) is placed into the system (in certain cases better data are obtained in the presence of 2.5 µg of SSB). The analysis begins by automatic dispensation of the substrate and enzyme reagents into the wells that contain the DNA samples. Nucleotides are dispensed sequentially, every 65 sec. The integrated mixer table ensures adequate mixing of the reagents during the analysis. The lens-array (placed under the reaction plate) focuses the light onto the CCD camera, which detects and transforms the light into signals which are displayed as peaks in the resulting Pyrogram. The height of the signal is proportional to the number of bases that have been incorporated. Within a Pyrogram the A signal is generally a little higher than the other peaks due to the use of a modified nucleotide, dATPα(S). The offset is generally 10 to 20% and the software compensates for this effect. All 96 wells are monitored simultaneously and any well may be selected during the run to monitor the sequencing process in real time. When the run is completed, the software automatically analyzes the resulting quantitative data.

III. RESULTS

A. CHOICE OF PCR PRIMERS

It is possible to use a PCR primer as sequencing primer, but in order to maximize the specificity and accuracy of the data a nested sequencing primer is recommended. To maintain the performance characteristics (the sensitivity and specificity of the system) in Pyrosequencing, samples should be sequenced when a clear and specific product band is seen on ethidium bromide-stained agarose gels after PCR amplification. Nested PCR is recommended for direct amplification of low copy numbers from cell lysates or in cases of unspecific amplification from DNA extracts. Generally, the shorter the PCR fragment the better, although fragments of up to 2000 bp have been tested with good results. The main limitation for using longer fragments is the yield from the PCR and the capacity of the streptavidin-coated beads (magnetic or Sepharose) to bind the fragment (the latter is not an issue if dsDNA is used as template).

It is advisable to design the sequencing primer before the PCR primers since the PCR primer to be biotinylated is determined by the strand to be analyzed. The strand complementary to the sequencing primer should be biotinylated at the 5′-end, i.e., for a forward sequencing primer the reverse PCR primer should be biotinylated and vice versa. If one of the procedures for dsDNA template preparation is used, there is no need for biotinylated PCR primers. For design of PCR and sequencing primers the OLIGO (http://www.oligo.net) software for primer design is recommended (software is also available from Pyrosequencing AB). Briefly, the PCR primers should typically be about 18 to 25 bases in length, of approximately the same GC-content as the fragment as a whole, and with approximately the same melting temperatures. The primers should preferably be more GC-rich in the 5′-end and less in the 3′-end for good specificity. They should not form
heavy hairpin loops or dimers with themselves or the other primers. Biotinylated PCR primer concentrations should be kept low (0.2 \( \mu M \)) to avoid interference with the Pyrosequencing assay.

B. DNA Preparation

1. Single-Stranded Template

If ssDNA is used as template, one of the PCR primers should be 5'-biotin labeled (for immobilization to streptavidin-coated magnetic or Sepharose beads), the other unlabeled. Prior to analysis, PCR products are converted to single-stranded templates onto which a sequencing primer is annealed. Both the immobilized biotinylated and nonbiotinylated strands in solution can be used as templates. Preparation of ssDNA template for Pyrosequencing can be performed manually or automatically using streptavidin-coated magnetic or Sepharose beads. Preparation of 96 samples in parallel can be performed manually using the multimagnet PSQ96 Sample Prep Tool (Pyrosequencing AB). Dedicated automated systems for ssDNA template preparation using magnetic beads are available from Magnetic Biosolutions AB (Stockholm, Sweden; http://www.magneticbiosolutions.se). By using the company’s new system for multiple regeneration of magnetic beads (the beads can be reused up to seven times without loss of capacity), the expense for magnetic beads can be dramatically reduced. In the new system, the DNA is released from the beads prior to sequencing, giving lower quenching and better kinetics.

If Sepharose beads are used, strand separation is performed in a filter plate by applying vacuum to remove the nonimmobilized DNA strand. To decrease the cost due to the use of biotinylated primers a strategy utilizing a general biotinylated primer has recently been described. In this strategy, a three-primer system is used for PCR with a target-specific primer containing a universal handle complementary to the general biotinylated primer (see Chapter 8 by Markus Schuelke for a similar strategy applied to fluorescent labeling).

2. Double-Stranded Template

Direct use of dsDNA instead of ssDNA has simplified the template preparation step prior Pyrosequencing. Two different methods for preparation of the dsDNA template have been described. In the first method, the excess amounts of primers and nucleotides after the PCR step are removed by enzymes. Apyrase and PPase very efficiently degrade nucleotides and PPi, and exonuclease I degrade the excess of PCR primers. The use of exonuclease I in combination with apyrase and PPase for sample preparation requires two steps prior to Pyrosequencing, and the low kcat of exonuclease I set the limit for how fast the procedure can be performed.

In the second method, two different types of oligonucleotides are used to prevent reannealing of remaining PCR primers to the template: oligonucleotides complementary to the PCR primers and 3'-end modified oligonucleotides with the same sequence as the PCR primers. Advantages with this strategy are (1) fast and simple template preparation procedure (one step), (2) no need for exonuclease I treatment, and (3) less problem with unspecific priming from loop structures and dimers. By careful oligonucleotide design, and/or by addition of SSB, problems with unspecific sequence signals due to mispriming can be reduced.

C. Pyrosequencing

Design of sequencing primers for Pyrosequencing follows the same criteria as for the PCR primers except that the melting temperature of this primer may be lowered. For optimal sequencing primer design, the design software from Pyrosequencing AB is recommended. As the sequencing reaction is run at 28°C, it is crucial to check the primer for self-annealing, especially at the 3’ end.

The sequencing procedure is very simple and performed automatically in the bench-top PSQ™96MA System or in the high-throughput PTP™ System utilizing proprietary software and
reagent kits from Pyrosequencing AB. A reagent cartridge is filled with six different reagents (enzyme mixture, substrate mixture, and four different deoxynucleotides) and inserted into the dispenser. A 96-well (or 384-well) plate containing the DNA samples annealed to sequencing primers is placed into the system. The sequencing process is started from the computer after the desired software has been chosen. The software enables individual control of the dispensing order for each well.

In Figure 18.2, sequencing data obtained from both (A) dsDNA and (B) ssDNA are shown. Accurate data for more than 30 bases were generated on both templates.

D. Applications

Pyrosequencing has opened new possibilities for performing sequence-based DNA analysis. The technique has been successful for both confirmatory sequencing and de novo sequencing. Pyrosequencing has been employed for many different applications such as genotyping,\(^5\) mutation detection,\(^12,13\) tag sequencing of a selected cDNA library,\(^14\) allele frequency quantification,\(^19,20\) multiplex analysis,\(^22\) monitoring resistance to HIV inhibitors,\(^23\) forensic identification,\(^24\) evolutionary studies,\(^29\)
bacteria typing,\textsuperscript{9,26} virus typing,\textsuperscript{10,22} fungal typing (unpublished data), sequencing of disease-associated genes,\textsuperscript{12} analysis of DNA methylation profiles,\textsuperscript{27} and sequencing of difficult secondary DNA structure.\textsuperscript{11} The latter application is especially important as structured DNA tracts (which appear, on average, approximately once per 2 kb in human cDNA) are not easily sequenced by conventional techniques due to compressions during gel electrophoresis. In addition, the Pyrosequencing technology enables determination of the phase of SNPs when the SNPs are near each other, allowing the detection of haplotypes.\textsuperscript{13,28} As the nucleotide delivery in Pyrosequencing can be specified according to the order of the sequence, longer and faster reads can be obtained, which makes the method especially well suited for resequencing applications such as clone checking and mutation scanning.

For analysis of SNPs, the 3’ end of a primer is designed to hybridize one or a few bases before the polymorphic position. In a single reaction, all the different variations can be determined. An SNP can easily be scored manually by comparison of predicted SNP patterns or automatically by pattern recognition software. Software programs for SNP and allele frequency studies as well as for tag sequencing are available. The assignment of quality values is based on a number of different parameters, including difference in match between the best and next best choice of genotypes, agreement between expected and obtained sequence around the SNP, signal-to-noise ratios, variance in peak heights, and peak width. The quality value is displayed as a color code.

IV. DISCUSSION

Pyrosequencing is a robust and easy to handle DNA sequencing technique. By using the commercial system available from Pyrosequencing AB, both the sequencing step and the data analysis step are performed automatically. The Pyrosequencing reactions are easy to set up, which makes the technique accessible for personnel other than specialized researchers.

Often, good sequencing data can be obtained regardless of the method used for template preparation. Where the sequencing data are of lower quality, the main reasons are usually (1) primer-dimers produced in the PCR, (2) primer-dimers produced by the sequencing primer, and (3) unspecific priming. Primer-dimers and unspecific priming can be avoided by careful primer design, the application of stringent conditions, and the use of “hotstart” PCR. However, if high flexibility in the primer design is desired, it is difficult to avoid the problem with primer-dimers and unspecific priming. In such cases, addition of SSB to the sequencing reaction helps.\textsuperscript{15,16,29} To further improve the sequence data quality and the specificity of sequence data, a sequencing primer distinct from the PCR primers can be used. Preferably, only PCR products that produced a clear and specific band on ethidium bromide-stained agarose gel should be sequenced. We recommend nested PCR for direct amplification from cell lysate or in case of unspecific amplification from the DNA extract.

Recent improvement of the Pyrosequencing chemistry has allowed sequencing of up to hundred bases,\textsuperscript{4} pointing to new applications for the technique in the field of long sequencing. The method will be especially suited for microbial typing, expression studies, mutation scanning, and clone-checking. In a recent validation of the Pyrosequencing technique against the 5’ nuclease (TaqMan) assay, identical results were obtained.\textsuperscript{30} It was concluded that Pyrosequencing is a robust and easy to handle method for large-scale SNP analyses. The multibase reading capacity of Pyrosequencing facilitated the positioning of the sequencing primer. In addition, the technology is both time- and cost-competitive when compared with the existing sequencing and genotyping methods.\textsuperscript{17,18,22,30,31}

ACKNOWLEDGMENTS

This work was supported by grants from the Swedish Research Council for Engineering Sciences (TFR) and the Swedish National Board for Industrial and Technical Development (NUTEK).
REFERENCES

I. INTRODUCTION

The general approach to DNA sequencing requires breaking a DNA molecule of interest into overlapping fragments, determining the order of bases in these fragments, and reassembling the obtained sequences to produce the sequence of the starting DNA molecule. This process generates a population of DNA molecules that share the same 5' end, but different 3' ends due to the incorporation of a particular chain-terminating dideoxynucleotide. The resulting set of molecules can then be separated by electrophoresis on denaturing polyacrylamide gels or capillary systems, forming a DNA “ladder” that can be followed for several hundred nucleotides. The traditional template-directed primer-walking strategies, however, require designing and synthesizing sequence-specific primers (typically 15 to 25 bases long) for a project and can typically be used only with that template.

We have developed a primer-directed DNA sequencing method (octamer sequencing technology [OST]) in which individual octamers are selected from a presynthesized primer library and used to prime multiple DNA sequencing reactions in parallel. The availability of such a library eliminates delays associated with designing and synthesizing gene-specific primers. Furthermore, the primers in our working octamer library can be used to sequence virtually any template. A computer program, electronic OST (eOST), was developed to identify the optimal octamer-primers for a particular template.
and to facilitate the process of primer selection. In addition, we discovered that nucleotide supplementation of octamer-primed sequencing reactions leads to an increase in the success rate of fluorescent sequencing and improves the quality of the sequence data obtained from difficult template regions.

II. MATERIALS AND METHODS

A. PRIMER SELECTION AND SYNTHESIS

Octamers used in various sequencing projects were chosen using the eOST analytical software. eOST integrates the base-calling software PHRED and allows identification of high quality template regions. The eOST schema is presented in Figure 19.1.

Library primers used in octamer-primed DNA sequencing reactions were synthesized by either Genosys Biotech, Inc., or Operon Technologies, Inc.

B. PURIFICATION OF PLASMID DNA TEMPLATES FOR OCTAMER-PRIMED SEQUENCING

To obtain plasmid DNA for sequencing, DH5α or XL1-Blue Eschericia coli cells were transformed using 4 ng plasmid DNA, plated on LB agar plates containing the appropriate antibiotic, and

![Diagram of the eOST strategy](www.taq.ir)
incubated at 37°C overnight. Single colonies were picked from the plates and used to inoculate 5 ml antibiotic-containing LB broth and grown overnight at 37°C. To further amplify the plasmid DNA, 20–50 µl of the 5-ml culture was transferred into 25–50 ml LB containing antibiotic and grown at 37°C with constant shaking overnight. Plasmid DNA was extracted using a QIAGEN Midi Kit (QIAGEN, Inc.) following the manufacturer’s instructions.

C. Octamer-Primed DNA Sequencing

1. Isotopic Sequencing

Components of the sequencing reaction:

25 pmol octamer primer
250 to 500 ng plasmid template
10 µCi of α-32P-labeled dATP
SequiTherm™ DNA Polymerase
Termination mix
Reaction buffer

In general, these reactions were performed as recommended by the manufacturers of the Sequi-Therm™ Cycle Sequencing Kit (EpiCentre Technologies, Inc.). To optimize octamer extension, the amount of primer was increased to 25 pmol per reaction, and an additional 1-min annealing step was introduced during cycling.

The sequencing reactions, assembled as described above, were cycled in a Perkin–Elmer GeneAmp® 9600 as follows:

Step 1 95°C 5 min (initial denaturation)
Step 2 95°C 30 sec (denaturation)
Step 3 30 to 40°C 1 min (annealing)
Step 4 70°C 1 min (extension)
Step 5 Repeat from step 2 above 29 times
Step 6 4°C Continuous (storage temperature)

2. Fluorescent Sequencing

Standard (nonsupplemented) sequencing reactions:

25 to 50 pmol octamer primer
200 ng plasmid template
2 µl of ABI PRISM BigDye™-Terminator Ready Reaction Mix (original version)
Reaction buffer

Reactions supplemented with dGTP and dTTP:

15 pmol octamer primer
200 ng plasmid template
2µl ofABI PRISM BigDye™-Terminator Ready Reaction Mix (original version)
Reaction buffer
75 µM dGTP (from freshly diluted stock solution)
5 µM dTTP (from freshly diluted stock solution)
Both types of fluorescent sequencing reactions were performed in a final volume of 10 μl. The sequencing reactions, assembled as described above, were cycled in a Perkin–Elmer GeneAmp® 9600 or MJResearch, Inc. PTC-200 thermocycler, as follows:

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>96°C</td>
<td>2 min (initial denaturation)</td>
</tr>
<tr>
<td>2</td>
<td>96°C</td>
<td>10 sec (denaturation)</td>
</tr>
<tr>
<td>3</td>
<td>40°C</td>
<td>1 min (annealing)</td>
</tr>
<tr>
<td>4</td>
<td>60°C</td>
<td>4 min (extension)</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Repeat from step 2 above 99 times</td>
</tr>
<tr>
<td>6</td>
<td>4°C</td>
<td>Continuous (storage temperature)</td>
</tr>
</tbody>
</table>

D. PREPARATION OF SEQUENCING EXTENSION PRODUCTS FOR ANALYSIS

The sequencing reaction extension products were precipitated using the standard sodium acetate/ethanol precipitation method. Immediately prior to their loading on a sequencing gel, the precipitated sequencing reactions were resuspended in 3 or 8 μl of loading buffer for fluorescent and radioisotopic reactions, respectively (loading buffer composition: 5 parts deionized formamide:1 part 25 mM EDTA (pH 8.0) containing 50 mg/ml blue dextran).

The samples were heated at 96°C for 2 min and 1.5 μl of the reaction was loaded onto a sequencing gel.

E. ANALYSIS OF DNA SEQUENCE READ-LENGTH AND DATA ACCURACY

1. Isotopic Sequencing

Sequencing gels were placed for 2 h against a BAS-IIIS imaging plate. The imaging plate was scanned using a Fuji BAS1000 Phosphoimager (Fuji Medical Systems USA, Inc.) and the data were analyzed using ImageGauge™ (Fuji Photo Film Corp., Ltd.) Software. The DNA sequence was manually determined.

2. Fluorescent Sequencing

The sequence information generated in octamer sequencing reactions was collected in 2X mode by an ABI PRISM™ 377 DNA Sequencer. The data were trimmed automatically using Sequencher, version 3.0 or 3.1 (GeneCodes, Inc.) with the following trimming parameters:

1. The sequences were trimmed until the first 25 bases at the 3′ and 5′ termini contained fewer than 3 ambiguities.
2. The leading and trailing ambiguous bases were removed.
3. All the sequence reads obtained by an individual primer in regular or supplemented sequencing reactions were aligned in a contig and trimmed beginning at a common starting point. In this analysis, data accuracy was determined using Sequencher, version 3.0 or 3.1 (GeneCodes, Inc.).

III. RESULTS

A. DEVELOPMENT OF THE OCTAMER-PRIMER LIBRARY

When developing a short-oligonucleotide library, two main criteria — length and priming specificity — need to be taken into account. Octamers are the shortest length primers that produce robust sequencing data and provide sufficient specificity in cycle sequencing. The size of a complete octamer library is 4⁸ or 65,536 primers. To be useful, a working primer library should contain
Octamer-Primed Cycle Sequencing

~1000 primers that can be used to sequence a broad variety of templates. The octamer library was initially optimized for use with radioisotopic sequencing. The following criteria were used to identify library octamers: (1) only octamers that have 50% GC content were included; (2) octamers were optimized with respect to amino acid and codon usage frequencies. Positions 1 to 3 and 4 to 6 were considered as adjacent codons, and positions 7 and 8 as the first two bases of a third codon; (3) Octamers containing either less frequently used codons or termination codons were excluded to reduce the size of the library to 4717 octamers. Among these, 958 complementary octamer pairs were chosen (allowing for double-strand sequencing) and can be used to sequence both genomic and cDNA templates. This number was further reduced to 566 complementary primer pairs by subtracting those octamers that were present in cloning vectors and primers that occur infrequently or exist in known repeats for the target organism.

To adapt the technology for fluorescent cycle sequencing chemistry, additional optimization of the octamer library was performed. The GC content of library octamers was increased to 75% and a new 970-member library was created, using the library design criteria. Also optimized were the annealing temperature, the number of reaction cycles, and the amount of octamer used to prime the reactions. We discovered that 40°C annealing temperature, 99 cycles, and 25 to 50 pmol of primer are optimal for each octamer to produce high quality sequence data (Figure 19.2).

B. Primer Identification Software

Presynthesized libraries are designed to make the process of primer selection for a particular project faster and easier. However, such libraries are often comprised of a large number of primers (approximately 1000 primers are in our working octamer library), which can delay this process. To make OST more time-efficient and user-friendly we introduced a computer program, eOST, into the process of primer selection. eOST integrates the base-calling software PHRED to identify high quality template regions and eliminate low quality, unreliable sequences.

Confidence in the template sequence is critical since OST conditions are optimized such that only perfectly matched octamers prime reactions. PHRED quality analysis software automatically identifies high-quality sequence prior to octamer-primer selection. The default quality cutoff value integrated into eOST is a PHRED value of 20. This value corresponds to a 1% chance that a certain base-call is not correct. Bases with quality below this cutoff value are considered to be of low quality. Regions containing high quality sequence are designated as "good" or suitable for octamer selection. These regions are then analyzed for matches with primers in the library using alignment settings with 100% match and an 8-base overlap.

An electronic version of the primer library is screened for the occurrence of individual primers, the strand location, and the frequency of occurrence within the known template sequence (electronic version available at http://hgsc.bcm.tmc.edu/~gmei/susan/). In the next step, the eOST program analyzes the template regions surrounding the octamer-binding site for potential secondary structures that might inhibit the efficiency with which the octamer-primer can initiate a sequencing.
reaction. The 5' terminus of the DNA fragment where the primer is located is folded toward the 3' terminus (the 5' and 3' template termini are defined relative to the primer position in the sequenced region; see Figure 3 in Reference 4). By increasing the amount of overlap by a single base, all possible secondary structures are generated. The eOST software calculates the stabilities (in Gibbs free energy — $\Delta G$) for all identified secondary structures, which are then sorted by their $\Delta G$ value. The lowest free energy value is considered to correspond to the most stable structure, and these areas can be avoided. Octamers are chosen to prime sequencing reactions only if the template secondary structures in the surrounding region have a $\Delta G$ higher than $-5.0$ kcal/mol.

C. OPTIMIZATION OF THE SEQUENCING REACTION CHEMISTRY

A potential problem associated with the use of presynthesized short-oligomer libraries is their reduced sequencing success rate compared to sequencing protocols using traditional length primers. Use of modified oligonucleotides or optimization of the sequencing reaction conditions have been viewed as tools to increase the success rate of short-oligo primed sequencing reactions.

OST was initially developed and optimized for use with isotopic sequencing protocols. This technique was later adapted for fluorescent, dye-terminator cycle sequencing. Both rhodamine and the subsequently introduced BigDye fluorescent chemistries can be used in octamer-primed sequencing reactions. However, BigDye octamer sequencing has a reduced success rate in comparison to rhodamine chemistry.

The two fluorescent chemistries differ with regard to their deoxynucleotide (dNTP) composition. Rhodamine chemistry contains dATP, dCTP, dTTP and, to minimize band compression, dITP in place of dGTP. In addition, BigDye chemistry contains dUTP instead of dTTP. This nucleotide was introduced to obtain more consistent peak pattern. Since both fluorescent chemistries have lower success rate than isotopic octamer sequencing, we viewed their altered dNTP composition as a potential cause of this problem. Our next step into the optimization of OST was to investigate and eliminate reasons for this reduced success rate.

Through primer extension experiments we discovered that the natural bases dGTP and dTTP are incorporated by the polymerase more efficiently than the nonnatural bases dITP and dUTP.
Thus, absence of dGTP and dTTP in the sequencing reaction mix can lead to reduced extension and result in reduced success rate of OST. To address this problem we supplemented octamer-primed sequencing reactions with various concentrations of dGTP and dTTP. We discovered that 75 µM dGTP and 5 µM dTTP increases the sequencing success rate and improves the quality of the sequence information (Figure 19.3). Out of 16 primers that failed to produce sequencing data of good quality in regular (nonsupplemented) reactions, 11 were able to prime in cycle sequencing after supplementation, and 9 of these generated accurate sequence reads of 500 or more bases.

IV. DISCUSSION

To overcome time, cost, and labor constraints associated with the design and synthesis of sequence-specific primers in traditional primer-directed sequencing, we proposed using a presynthesized library of carefully selected octamer-primers. These octamers are used to sequence a variety of different templates without the need to design additional template-specific primers. The availability of an octamer-primer library accelerates the process of DNA sequencing due to the immediate access to the next sequencing primer and the ability to process reactions in parallel. Thus, the main advantages resulting from a primer library sequencing approach are that primer identification and choice are more time and cost efficient.

The working octamer library consists of approximately 1000 primers and can be used to sequence essentially any template. These octamers are optimized with regards to GC content, amino acid occurrence, and codon usage frequencies. Unlike gene-specific primers, all library primers have a unified cycling regimen due to their similar nucleotide composition. This allows different DNA samples to be processed and sequenced in parallel, which further reduces time-related costs.

The final improvements to OST — octamer selection through eOST analysis and increased extension efficiency through dGTP/dTTP supplementation — bring the success rate of fluorescent sequencing to ~89%. Isotopic chemistry success rate is predicted to be essentially 100% because gel exposure times may be varied to compensate for differences in signal strength.

ACKNOWLEDGMENTS

We would like to thank Mitsu Reddy and Nathan Agnew for helpful suggestions. This work was supported by the NIH (grant R29-HG01151).

REFERENCES


20 In Situ Amplification of cDNA

Gerard J. Nuovo

I. INTRODUCTION

The ability to isolate, amplify, and characterize specific DNA and RNA sequences has led to tremendous changes in such diverse fields as basic research, forensic pathology, and the study of infectious disease. In many instances, it is highly desirable to determine what specific cell type contains the DNA or RNA sequence of interest. One example, discussed in this chapter, are the specific targets of infection by the virus HIV-1. In theory, this question could be addressed by in situ hybridization. However, this method requires the presence of at least 10 target sequences in an intact cell in order to produce a detectable signal. In many instances only one or a few targets may be present in a given cell. HIV-1 is a classic example, where one copy of the DNA provirus of HIV-1 can integrate into the nucleus of the host cell and remain latent for many years. Solution phase PCR, on the other hand, can detect as few as 1 target per 100,000 cells, but information about its localization is lost due to the obligatory destruction of cells for DNA extraction. By combining the cell localizing ability of in situ hybridization with the high sensitivity of PCR, one can readily and routinely detect 1 to 10 target sequences in a cell and — either by using its cytologic features or colocalizing probes — determine the exact nature of that cell. Thus, PCR in situ hybridization (for DNA) and in situ reverse transcriptase (RT) PCR (for RNA) are the methods of choice for detecting low copy sequences in intact tissue samples.
II. EQUIPMENT AND REAGENTS

A. INSTRUMENTATION

An important decision one needs to make is whether to use an instrument dedicated to in situ PCR or adapt a standard solution phase PCR machine for this purpose.

Machines dedicated to in situ PCR are made by several companies, all of which employ a flat aluminum surface on which the slides are placed directly. Perhaps the most commonly used machine is from Perkin–Elmer, allowing 10 slides to run at a given time. It uses plastic coverslips called Amplicovers that are held down by aluminum clips (Ampliclips) in order to keep the PCR reagent over the tissue during the amplification process. This author has a great deal of experience with this method which requires minimal hands-on work. Other machines that are dedicated to in situ PCR do not use coverslips and clips to hold the reagents over the tissue but rather some type of sealant. One example of a sealant machine is from MJ Research and employs twin-heated towers. The spaces between slides are very thin and require the use of a special solution called SelfSeal which is mixed with the amplifying solution and prevents its evaporation during the amplification process. The SelfSeal reagent is very useful and can be employed in any of the other machines.

Most of the other machines dedicated to solution-phase PCR can be adapted for in situ PCR. Importantly, all require a method to keep the amplifying solution over the slide, for example, by using the SelfSeal reagent. Another method widely employed by this author involves aluminum boats. Here, the amplifying solution is placed over the tissue which is then covered by a polypropylene coverslip cut to size and anchored on the slide with a few small drops of nail polish. Finally, adding mineral oil to cover the slides prevents the evaporation of reagents during the cycling process. An advantage of this method is that it allows one to perform amplifications on two to three tissue sections on a single silane-coated slide.

B. REAGENTS

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silane-coated glass slides</td>
<td>Fisher Scientific (Pittsburg, PA)</td>
</tr>
<tr>
<td>Proteinase K</td>
<td>Enzo Biochemistry (Farmingdale, NY)</td>
</tr>
<tr>
<td>Pepsin</td>
<td>DAKO Laboratory (Carpenteria, CA)</td>
</tr>
<tr>
<td>RNase-free water</td>
<td>Fisher Scientific</td>
</tr>
<tr>
<td>RNase-free DNase</td>
<td>Boehringer Mannheim (Indianopolis, IN)</td>
</tr>
<tr>
<td>In situ PCR reagent pack</td>
<td>Perkin–Elmer/Applied Biosystems (Foster City, CA)</td>
</tr>
<tr>
<td>(includes taq polymerase and its buffer,</td>
<td></td>
</tr>
<tr>
<td>dNTPs, MgCl2)</td>
<td></td>
</tr>
<tr>
<td>In situ RT reagent pack</td>
<td>Perkin–Elmer</td>
</tr>
<tr>
<td>(includes EZ rTth polymerase and its buffer,</td>
<td></td>
</tr>
<tr>
<td>dNTPs, MnAcetate)</td>
<td></td>
</tr>
<tr>
<td>Bovine serum albumin</td>
<td>Fisher Scientific</td>
</tr>
<tr>
<td>Digoxigenin dUTP</td>
<td>Enzo Biochemistry</td>
</tr>
<tr>
<td>Stringency wash buffer</td>
<td>Enzo Biochemistry</td>
</tr>
<tr>
<td>Detection reagent buffer</td>
<td>BioGenex (San Ramon, CA)</td>
</tr>
<tr>
<td>Antidigoxigenin alkaline phosphatase conjugate</td>
<td>Enzo Biochemistry</td>
</tr>
<tr>
<td>NBT/BCIP</td>
<td>Enzo Biochemistry</td>
</tr>
<tr>
<td>Nuclear fast red</td>
<td>Enzo Biochemistry</td>
</tr>
</tbody>
</table>
III. EXPERIMENTAL PROTOCOLS

A. STARTUP

The purpose of the startup protocol\textsuperscript{1,2} is to determine the optimal conditions for protease digestion of the sample being investigated. This is necessary, firstly, because DNA and RNA molecules form bonds with the cellular proteins when tissue is fixed in formalin or a similar crosslinking fixative. These protein–nucleic acid and protein–protein crosslinks provide “support” to the nuclear and cytoplasmic scaffolding (which is important for good cytologic detail) and inactive degradative enzymes such as RNases, but may also impede reverse transcription. Second, after the tissue is fixed in formalin it must be embedded in paraffin wax which requires high temperatures. This invariably leads to nicks in the genomic DNA which can become starting points for DNA synthesis during the PCR step of \textit{in situ} RT-PCR.\textsuperscript{1,2} Complete digestion of genomic DNA with DNase prior to the \textit{in situ} RT and PCR steps is therefore critical but depends, in turn, on the removal of protein-DNA crosslinks by protease digestion. Thus, the optimum digestion time in the protease solution will prevent nonspecific DNA synthesis from the genomic DNA while still permitting excellent cytologic detail. I would like to stress that optimal protease and DNAse pretreatment, determined with the startup protocol, is critical and generally sufficient for plant as well as animal tissue, in either case allowing detection of bacteria, viruses, or mRNA.

There is a good correlation between the length of time a tissue is fixed in formalin and the optimal protease digestion time for \textit{in situ} RT-PCR.\textsuperscript{4} However, for most biopsies one does not know how long the tissue was fixed and this optimal time is also dependent on tissue type, being less for noncohesive cells that have relatively low protein concentrations in their cytoplasm (e.g., lymphocytes in lymph nodes) and greater for tissues like skeletal muscle and skin. For surgical biopsies and autopsies, I recommend the following startup protocol:

1. Place three sections (4 microns each) on a silane-coated glass slide.
2. Deparaffinize the tissue.
3. Prepare 2 mg/ml pepsin (20 mg of pepsin + 9.5 ml water + 0.5 ml of 2 NHCl) — either use immediately, or freeze in 1-ml aliquots, and use immediately after thawing.
4. Digest the three sections in pepsin for 15, 30, and 60 min.
5. Wash in water, then 100% ethanol, and air dry.
6. Add RT/PCR reagent (10 \(\mu\)l of \(rTth\) buffer, 1.6 \(\mu\)l each of dATP, dCTP, dGTP, dTTP, and 2% BSA, 2 \(\mu\)l of \(rTth\), 12.4 \(\mu\)l of 10 \(mM\) Mn acetate, 0.6 \(\mu\)l of 1 \(mM\) digoxigenin dUTP, 17.0 \(\mu\)l of sterile water).
7. Incubate at 60\(^\circ\)C for 1 h.
8. Remove mineral oil (if applicable) in xylene then 100% ethanol, wash in stringency buffer (0.2 X SSC and 2% BSA at 60\(^\circ\)C for 10 min).
9. Incubate with antidigoxigenin-alkaline phosphatase conjugate (1:150 dilution) for 30 min.
10. Detect using NBT/BCIP = nitroblue tetrazolium and bromochloroindolyl phosphate prepare fresh by adding 35 \(\mu\)l of each into 15 ml of pH 9.5 detection buffer (BioGenex) that has been prewarmed to 37\(^\circ\)C.
11. Counterstain in nuclear fast red for 5 min.
12. Coverslip in Permount through xylene.

Color Figure 20.1 shows an example of a startup experiment.*

* See color figures following page 234.
FIGURE 20.1 Determining optimal protease digestion with the startup protocol. A section of brain tissue was digested in pepsin for (A) 15 min or (B) 60 min, and treated as described in the text. Note the lack of detectable DNA synthesis in panel A and the strong signal in the nuclei in panel B. (C) With the longer protease digestion, the signal is completely lost with overnight digestion with RNase-free DNase. (D) However, as expected, background is still evident after DNase digestion if 15 min of protease digestion is used. (See color version of this figure following page 234.)
B. *In situ* RT–PCR

1. Digest the tissue/cell sample with pepsin using the optimal digestion time, determined as described above.
2. Remove the protease with water and 100% ethanol baths, then air dry.
3. DNase two of the three tissue sections overnight at 37°C (use 10 ml per slide — prepare the solution by adding 1 µl of the RNase-free DNase, 1 µl of rTth buffer, and 8 µl of RNase-free water). The third section should be incubated in buffer solution only (i.e., 1 µl of the rTth buffer plus 9 µl of RNase-free water).
4. Remove the DNase with water and 100% ethanol brief washes, then air dry.
5. This is the key step of *in situ* RT–PCR. It is important to realize that one should perform three different experiments on a single slide:
   - *The negative control section* (which has been treated with DNase) gets the *in situ* RT–PCR solution either with or without irrelevant primers. The definition of irrelevant primers is that the target they represent could not possibly be present in the tissue being analyzed. I routinely use either rabies- or hantavirus-specific primers as these viruses are not going to be present in most samples. I prefer to use irrelevant primers rather than no primers, as this also serves as a check for nonspecific DNA synthesis from mispriming due to inadequate DNase digestion.
   - *The positive control section* has not been treated in DNase and thus should show robust nonspecific DNA synthesis, regardless of the primers. One can use the *in situ* RT–PCR solution with the specific primers of interest for the convenience of making the same solution for the positive control and for the test slide.
   - *The test section* has been treated with DNase and is incubated with the *in situ* RT–PCR solution which, of course, contains the target-specific primers of interest.
6. Prepare the *in situ* RT–PCR solution by adding 10 µl of rTth buffer, 1.6 µl each of dATP, dCTP, dGTP, dTTP, and 2% BSA, 2 µl of rTth, 12.4 µl of 10 mM Mn acetate, 0.6 µl of 1 mM digoxigenin dUTP, 3 µl of primers (stock solution at 20 µM), 14.0 µl of sterile water. The use of the enzyme rTth, which can perform both RT and PCR functions, simplifies the technique by eliminating the need for two separate steps (RT and PCR) with different reagent requirements. Add to the slide and incubate at 65°C for 30 min.
7. Denature at 95°C for 5 min and amplify the cDNA by 20 cycles of 60°C for 90 sec and 95°C for 45 sec.
8. Follow steps 8 to 12 of the startup protocol (see above).

C. **INTERPRETATION OF THE DATA**

The importance of having a basic knowledge of surgical pathology for the interpretation of *in situ* data — whether immunohistochemistry, *in situ* hybridization, or *in situ* PCR — cannot be stressed enough. For simplicity’s sake, let us call the target-specific signal just “signal” and use the word “background” for any nonspecific staining of the same color. Background is just as common with *in situ* hybridization and immunohistochemistry as it is with *in situ* RT–PCR. The keys to recognizing it are:

1. **Controls**

This is simple if one does the controls recommended above on the same glass slide as the test was done. A successful *in situ* RT–PCR run is defined by no background with the negative control (DNase with irrelevant primers) and an intense nuclear-based signal in at least 50% of the cells with the positive control (no DNase) in different cell types. It follows that the definition of an unsuccessful *in situ* RT–PCR run is also simple; background, nuclear-based colorization is evident with the negative control (DNase with irrelevant primers). When this occurs, one typically sees...
either a weak or no “signal” with the positive control (no DNase) but, for ease of interpretation it is simpler to focus one’s attention on the negative control (DNase with irrelevant primers).

2. Histological Pattern of the Signal

A good surgical pathologist will appreciate the value of this tool. If one is looking for a lymphocyte marker and does immunohistochemistry for such, and then sees strong “signal” in lymphocytes and squamous cells, then by definition there is background, and the experiment must be redone. This is particularly useful when studying viruses, as their specific cell tropism is usually well known (Color Figure 20.2). Of course, there will be instances where the specific histological distribution of a given target will not be known and one must rely more on the correct interpretation of the controls as detailed above. Still, one should always be suspicious of background when a “signal” is seen in many diverse cell types with a given probe, primer set, or antibody.

IV. DISCUSSION

A. Common Pitfalls and Troubleshooting Tips

There is one overriding and all-important pitfall that can create problems — or more specifically background — with in situ RT–PCR. This is inadequate protease digestion. Unless there is adequate protease digestion, the DNase will not be able to have access to the entire genomic DNA. The end result will be relatively long segments of double-stranded DNA that will have nicks (from either the incomplete DNase digestion or the heating required during the paraffin embedding). The rTth polymerase will avidly use these nicks to synthesize new DNA into which the digoxigenin dUTP will be incorporated; the end result will be nonspecific background occurring in the nucleus of the different cell types present in the tissue. The negative control (DNase with irrelevant primers) is the best way to determine whether the protease digestion is adequate; a nuclear-based background in any cells will tell the investigator that the protease digestion was not sufficient. Thus the experiment must be repeated, this time increasing the protease digestion time.

The reverse common mistake is over-digestion. This is easy to recognize; the nuclei of the cells are not evident after counterstaining with the nuclear fast red. Rather, the fibers between cells (called reticulin) tend to persist, and one sees a maze-like arrangement of these fibers without any cells. The signal will be completely lost and the experiment must be redone, this time reducing the protease digestion time.

Using labeled primers instead of tagged nucleotides eliminates a key cause of background — namely, nonspecific incorporation of the tagged nucleotide during DNA synthesis. However, conventionally labeled primers often do not yield as strong a signal. Universal energy transfer-labeled primers allow more robust RNA detection in situ.15

B. Budget, Expertise, and Effort Required

The most expensive single item is the thermal cycler. As noted at the beginning of this chapter, one must decide whether to purchase a cycler dedicated to in situ PCR or one that is made for solution phase PCR and then modified for in situ applications. Clearly, the latter is the less expensive route.

With regard to expertise, one cannot stress enough the importance of becoming proficient in standard in situ hybridization and immuno-histochemistry before one attempts in situ RT–PCR! This author has seen people with a lot of experience with in situ hybridization master in situ RT–PCR in as short a time as a few days. This has never been my experience with people who have done little to no in situ hybridization. Clearly, this is related to the effort required. If one has extensive experience doing in situ hybridization, little extra effort should be required to perform successful in situ RT–PCR.
FIGURE 20.2 *In situ* RT-PCR for RNA viruses. (A) Section of brain from a woman with encephalitis; RT-PCR with primers detecting Eastern equine encephalitis virus. Note the strong cytoplasm-based signal in the neuron after *in situ* RT-PCR. The tissue was digested with protease for 60 min of protease digestion; the negative and positive controls are those shown in Figure 20.1.

(B–D) Sections of spinal cord from a man with gastroenteritis and paralysis. (B) Intense signal with polio virus *in situ* RT-PCR. (C) Higher magnification — note how the signal localizes to the cytoplasm of the large anterior horn motor neurons. These neurons are well known targets of this virus, another indicator for the specificity of the reaction. (D) Negative control using primers specific for human papillomavirus instead of poliovirus primers. (See color version of this figure following page 234.)
C. OTHER USEFUL APPLICATIONS OF THE METHOD

This last section briefly highlights the utility of this method in a variety of research and diagnostic modalities.

First, an understanding of HIV-1 pathogenesis sparked the initial work with in situ RT–PCR and PCR in situ hybridization. These two methodologies demonstrated the following:

• Pre-AIDS is due to a massive, primarily latent infection by HIV-1; over 30% of the CD4 cells are infected prior to the onset of AIDS-defining symptoms.5–7
• Sexual transmission of HIV-1, which is the primary mode of spreading of the disease, is due to infection of endocervical macrophages in the cervix and anorectal junction and spermatogonia in the testes.8,9
• AIDS dementia is due to infected macrophages/microglia, astrocytes, and neurons with concomitant upregulation of several cytokines including MIP, TNF α, and iNOS in neighboring but noninfected cells.10–12

Second, in situ RT-PCR has aided in establishing the cause of death in cases where there was no other clearcut evidence:

• Stillborn infants and infants who die soon after birth after demonstrating severe respiratory problems often are infected with coxsackie virus.13
• Similarly, infants born with low APGAR scores who later develop CNS dysfunction are often infected with the coxsackie virus, acquired via infection of the placenta.13
• Unexpected, rapid deaths in otherwise healthy individuals are most often due to viral myocarditis, which includes coxsackie virus, rotavirus, influenza virus A and B, and HIV-1.14

Last, an important future application of PCR in situ hybridization could be the study of local gene expression and its regulation at the single-cell level. For example, we showed recently that silencing of p16 by hypermethylation of the promoter was a common event in squamous cell cancer oncogenesis.16

ACKNOWLEDGMENTS

The author greatly appreciates the financial support from the Lewis Foundation and by NIH grant NCI (RO1 HL-00-012) (GJN).

REFERENCES

In Situ Amplification of cDNA

Detection of a Single-Copy Sequence on Human Metaphase by Cyclic PRINS

Kirsten Mennicke, Sven Opitz, and Eberhard Schwinger

I. INTRODUCTION

Primed in situ labeling (PRINS) is not merely an alternative to conventional fluorescent in situ hybridization (FISH) but a complementary approach. By combining features of FISH with polymerase chain reaction, it provides high sensitivity and specificity for in situ target sequence detection. In the original protocol described by Koch et al., an unlabeled oligonucleotide primer is annealed to its complementary target sequence and extended by DNA polymerase in presence of labeled nucleotides. The extended fluorescent labeled target is then visualized by fluorescent microscopy. In contrast to standard “tube PCR,” however, the mean length of extension generated by PRINS is only approximately 40 nucleotides and the amount of incorporated label is relatively low. Consequently, PRINS was mainly applied for targets containing repetitive sequences and the identification of single-copy sequences was achieved only in very few cases during the first decade after its invention. Recently, however, several improvements of the PRINS technique have been described that allowed successful localization of single-copy genes such as DMD, SRY, and SOX3. Moreover, PRINS has been successfully applied in the screening for chromosomal microdeletions in Prader-Willi/Angelman and DiGeorge/velocardiofacial syndrome. Interestingly, various reports emphasize different modification steps of the original PRINS protocol. Hindkjaer et al. used a primer cocktail of 18 primers to detect a sequence of the p53 gene on chromosome 17p13. The disadvantage of this approach is the...
same as in other multiplex PCRs; primer selection and construction is as difficult since all must have the same annealing temperature. Troyer et al. used a cyclic PRINS approach for target amplification. However, technical limitations restrict this approach since most commercially available cover slips and glue do not withstand several cycles of annealing, extending, and denaturation. Kandandale and Tharapel et al. improved the sensitivity mainly by applying the Tyramide Signal Amplification System™ (NEN Life Science Products, Boston, MA) in the immunochemistry step of the PRINS protocol. We have introduced a new chamber system for in situ applications of PRINS on slides which allows overcoming of some of the main disadvantages of the original PRINS protocol: (1) reaction conditions remain constant during many cycles, guaranteeing reliable results, (2) the open chamber allows for hot start and nested in situ PCR which is a unique improvement in the development of PRINS, and (3) since the reaction mix can be removed after cyclic amplification, aliquots can be run on a gel to check for successful amplification of an internal control. Our system can be used for the detection of high and low copy repetitive sequences, and it worked well for a 273-bp sequence of the single-copy DMD gene.

II. EXAMPLE

A. INSTRUMENTATION

Coplin jars
PCR tubes 0.5 ml ultrathin (Biozym Diagnostics, Olgendorf, Germany)
Hot glue pistol (Pattex Supermatic, Henkel, Düsseldorf, Germany)
Silicon glue (Pattex hot sticks transparent, Henkel, Düsseldorf, Germany)
Water bath with humidified chamber
Thermal cycler with flat bed (Omnigene flatbed, Hybaid, Teddington, U.K.)
DNA gel-electrophoresis apparatus (optional for control of genomic amplification)
Phase contrast microscope (Zeiss axioplan, Jena, Germany)

B. REAGENTS

Dideoxynucleotide mixture: 2.5 mM each of ddATP, ddCTP, ddGTP, and ddTTP (Boehringer Mannheim, Mannheim, Germany)
Taq-DNA-polymerase (Boehringer Mannheim or Appligene Oncor, Heidelberg, Germany)
Taq Polymerase 10 x buffer: 500 mM KCl, 15 mM MgCl₂, 100 mM Tris HCl (pH 8.3)
Nucleotide mixture: 60 µl aliquots containing 1.5 µl of each 10 mM dATP, dCTP, dGTP; 1.0 µl of 10 mM dTTP (all Boehringer Mannheim); 4 µl of glycerol 80% (Merck, Darmstadt, Germany); 7.5 µl of 10 x Taq buffer and 43 µl of redistilled water
Primers (synthesized by MWG Biotechnology, Ebersberg, Germany):
DMD-Ex8-iiF = 5’-TCT GGA GGA CAT TCA TGG ACA ATT CAC TG-3’
DMD-Ex8-iiR = 5’-TTC TTT AGT CAC TTT AGG TGG CCT TGG CA-3’
Antidioxygenin-fluorescein, Fab fragments (Boehringer Mannheim, Mannheim, Germany)
Blocking solution (Boehringer Mannheim, Mannheim, Germany)
Formamide (Fluka, Deisenhofen, Germany)
Xylol (Merck, Darmstadt, Germany)
Ethanol 99%
DAPI and propidium iodide in antifade (Oncor Appligene, Heidelberg, Germany)
Stop buffer: 50 mM NaCl, 50 mM EDTA, pH 8
Washing buffer: 4 x SSC, pH 7.0, 0.05% Tween 20 (1 x SSC: 150 mM NaCl, 15 mM sodium citrate)
C. EXPERIMENTAL PROTOCOL

1. Standard Chromosome Spreads

Chromosomes were prepared from peripheral blood lymphocytes by standard procedures. Microscope slides were marked on the bottom side with a center-cross by diamond pen and subsequently precooled on a deep-freezer bag (–20°C). When a thin humid condensation layer appeared on the slide, 4 μl of methanol/acetic acid fixed-cell suspension was dropped onto the slide over the marking. Slides were air-dried, dehydrated using a series of increasing ethanol concentrations (70, 90, and 99%, 3 min each) and air-dried again. The slides could be stored at room temperature for 1 to 3 d.

2. Pretreatment with ddNTP

Single-strand nicks in the chromosomal DNA can act as nonspecific initiation points in in situ PCR (see also Chapter 20 by Gerard Nuovo on in situ RT-PCR). Obviously, DNAse digestion is not an option in PRINS, therefore slides were treated with ddNTP to seal the nicks. The following reaction mixture was prepared in an Eppendorf tube:

- 2.5 μl ddNTP mixture (1 mM)
- 2.5 μl 10 x Taq buffer
- 0.1 μl Taq polymerase (5U/μl)
- Redistilled water add 25 μl

The reaction mixture was applied to the slide and sealed with a cover slip. The slide was heated on a thermoblock at 72°C for 5 min. Subsequently, the slide was transferred to stop buffer for 1 min, washed in 2 x SSC for 10 min with one change after 5 min, passed through an ethanol series (70, 90, and 99%, 3 min each) and air-dried again.

3. Chemical Denaturation of Chromosomal DNA

The slides were incubated in 70% formamide; 2 x SSC for 2 min at 70°C and passed through an ice-cold (–20°C) ethanol series (70, 90, and 99%, 3 min each) and air-dried again. After this step slides could be stored at –20°C for several weeks.

4. Cyclic PRINS Protocol

In order to achieve optimal in situ DNA accessibility we developed a modified two-step cyclic PRINS protocol. The aim of the first cyclic protocol is only chain elongation and amplification of the target sequence. Since hapten labeled oligonucleotides cannot pass the local compact chromatin organization, this step was performed in the presence of unlabeled oligonucleotides. The reaction mix for the following cycles of target amplification contained dig-dUTP as labeled. NOTE: For the detection of low copy sequences, we achieved excellent results with a one-step cyclic PRINS protocol, applying dig-DTP already from the beginning. The two-step protocol was especially developed for small single-copy sequences:

1. For the first PRINS reaction, prepare the following reaction mixture in an Eppendorf tube:
   - 0.5 μg of each primer (DMD-Ex8iiF and DMD-Ex8-iiR)
   - 60 μl nucleotide mixture
   - Additional 0.5 μl 10 mM dTTP
   - Redistilled water ad 65 μl

2. Fill the reaction mixture into the reaction chamber, mount with 70 μl of paraffin oil, and heat the slide on the thermoblock at the annealing temperature of 64°C for 5 min.

www.taq.ir
3. Dilute 2 U Taq polymerase in 10 μl buffer and add to the reaction mixture after 1 min.

**Note:** Using a hotstart for cyclic PRINS prevents unspecific activity of the Taq polymerase.

4. Run the first temperature cycle program:
   - 64°C, 5 min; 72°C, 10 min (1st cycle)
   - 94°C, 4 min; 64°C, 3 min; 72°C, 6 min (2nd cycle)
   - 94°C, 1 min; 64°C, 1 min; 72°C, 3 min (3rd to 15th cycle)

After the last cycle the slides were held at 72°C for 5 min.

**Note:** The annealing temperature for the primer must be carefully optimized.

5. Add 5 μl 1mM dig-11-dUTP and 1.5 U Taq polymerase were added to the reaction mixture. Run the second temperature cycle program:
   - 94°C, 2 min; 64°C, 3 min; 72°C, 6 min (1st cycle)
   - 94°C, 1 min; 64°C, 1 min; 72°C, 3 min (2nd to 5th cycle).

**Note:** The relation of unlabeled to labeled nucleotides varies between reported PRINS protocols. We achieved the best results with the relation [dTTP]:[dig-dUTP] = 2:1.

6. Preheat stop buffer to the specific annealing temperature (here: 64°C). Mount plastic cover slips (10 x 10 cm) with a central 1-cm slit over the slide. Only the reaction chamber should penetrate the slit (Figure 21.1d). When 1 μl of stop buffer was applied in the reaction mixture, the paraffin oil was rinsed and could flow over the plastic cover slip without contacting the slide. After removal of the paraffin oil, both chamber and cover slip were quickly taken off and the slide transferred to a staining jar with stop buffer for 1 min.

7. Pass the slides through an ethanol series at room temperature (70, 90, and 99%, 3 min each). Remnants of glue can now be removed easily. Wash the slide in xylol for 3 min and pass it through a descending ethanol series (99, 90, and 70%, 3 min each).

8. Transfer the slides to a staining jar with a screw cap filled with washing buffer. Wash for at least 30 min with gentle agitation (change washing buffer after 15 min). Slides can be held at this stage overnight if convenient.

5. **Visualization of Labeled DNA**

1. Dilute 5 μl of antidigoxigenin-fluorescein Fab fragments in 1 ml of 1% blocking buffer. Incubate slides with the diluted antidigoxigenin fluorescein solution in a dark humidified chamber for 10 to 30 min.

2. Wash slides again 3 times for 5 min under gentle agitation. Mount in antifade solution containing either propidiumiodide or DAPI. All slides in our protocol were examined under a Zeiss Axioplan microscope fitted with a Chroma Technology triple bandpass filter (DAPI/FITC/Texas red) and a Hamatsu CCD camera.

III. **DISCUSSION**

As exemplified for a 273-bp sequence in exon 8A of the DMD gene, we describe a general method for the in situ detection of single-copy sequences in metaphase spreads. For routine detection of higher repetitive sequences, the original or slightly modified single-cycle PRINS protocol is efficient, presenting already a cheap, rapid, and simple alternative method for the detection of marker chromosomes in lymphocyte and amniocyte metaphase spreads and of numerical abnormalities in interphase nuclei, e.g., amniocytes, tumor cells, spermatozoa, and...
first polar bodies of oocytes. Moreover, the high specificity of already one PRINS cycle enables the visualization of only one base pair differences in repetitive sequences. However, up to now there are only few cytogenetic laboratories which use this method routinely. Despite the theoretical potential of this methodology there are four main problems which become apparent when it is practiced:
1. **Reliability of standard conditions:** In opposite to many commercial available FISH tests, there are only few systems available for routine PRINS. They are mostly limited to one cycle procedures and cannot be used for nested *in situ* PCR. Our system is cheap, and allows constant reaction conditions to be maintained throughout the entire temperature cycle program, resulting in high reproducibility and standardization of the PRINS protocols.

2. **Choice of primers:** First applications of PRINS with one pair of primers postulated that optimal target length should be a least 1000 bp.\(^9\) In our experience with the detection of low copy sequences, the amplification with two primers is linear rather than exponential, suggesting an incomplete amplification of the target. This even applies to very small target sequences of 170 bp.\(^6\) Thus, our advice is to use amplicon lengths between 150 to 300 bp. We designed our primers with the computer program OLIGO (Med-Probe, Oslo, Norway). Optimal concentration of primers vary between 0.5 and 1.6 mg for each single primer per 70 μl reaction mix and has to be adjusted for optimal signal to background ratio.

3. **Accessibility of condensed chromatin:** Both FISH and PRINS enable resolution of DNA sequences separated by a few megabases in metaphase or prometaphase chromosomes. The low end of resolution is approximately 20 kb in interphase nuclei by FISH. With PRINS, much smaller sequences can be detected. Potentially every known gene sequence can be used to design primers. However, the compact local chromatin structure of metaphase chromosomes limits amplification and results in lower product yields compared to naked DNA in a standard PCR. We improved *in situ* amplification by adding an initial chain elongation step with unlabeled nucleotides only. This step is advisable for sequences where no signal is obtained in a normal cyclic PRINS protocol.

4. **Sensitivity:** We achieved sufficient incorporation of labeled nucleotides in the later cycles of chain amplification to detect a single copy sequence by PRINS. While a similar sensitivity has been reported using an improved labeled nucleotide detection kit (Tyramide Signal Amplification System™, NEN Life Science Products, Boston, MA),\(^8\)-\(^10\) our protocol is cheaper for routine use. We were able to halve the costs for detection of a single-copy gene *in situ* by PRINS, in comparison to FISH. Our nested *in situ* PCR protocol, combined with an improved detection kit, might be a further milestone in the development of *in situ* visualization methods for very small sequences.

**ACKNOWLEDGMENTS**

We thank Dr. T. Harrer for performing the experiments described here and Dr. M. Schürmann for valuable advice in the construction of primer pairs. The authors thank Professor Stürzbecher for the kind assistance in preparing the manuscript.

**REFERENCES**


Nucleic Acid Quantitation
Meaningful Quantification of mRNA Using Real-Time PCR

Stephen A. Bustin

CONTENTS

I. Introduction ................................................................................................................. 225
II. Types of Quantitative PCR Assays .................................................................................. 227
   A. Relative Quantification ............................................................................................. 227
   B. Absolute Quantification ............................................................................................. 227
III. Critical Factors ............................................................................................................. 228
   A. Normalization ............................................................................................................. 228
   B. Quantification and Quality Assessment of RNA ....................................................... 229
   C. Reproducibility .......................................................................................................... 230
   D. Relevance ..................................................................................................................... 230
   E. Reverse Transcription ................................................................................................. 231
IV. Conclusions .................................................................................................................. 231
Acknowledgments .................................................................................................................. 231
References .................................................................................................................................... 232

I. INTRODUCTION

Reverse transcription (RT) followed by polymerase chain reaction (PCR) is the most sensitive method for the detection of low abundance mRNA and has become widely used for the quantification of steady-state mRNA levels. Assays are easy to perform, capable of high throughput, and can combine high sensitivity with reliable specificity. However, careful experimental design and validation remain essential for accurate quantitative measurements of mRNA levels. Two major types of assays exist, based on endpoint and real-time quantification of PCR product, respectively.

The main advantage of endpoint assays is that they are cheap, since the PCR step can be carried out on a conventional thermal cycler and the analysis is done using a standard gel electrophoresis setup. A second advantage is that amplicons can be distinguished on the basis of their size. Indeed, some applications, e.g., allelic discrimination or SNP assays, do not require real-time analyses at all. However, there are distinct disadvantages to conventional endpoint assays. They are significantly more labor-intensive and less capable of high-throughput as they require post-PCR processing, usually involving gels, plus additional probing or sequencing to confirm the identity of any amplicon generated. The dynamic range of a gel scan is much less than that of any real-time system and quantification is really only semiquantitative. However, endpoint analysis carried out using fluorescent reporters and real-time instruments are the method of choice for SNP analyses.

Real-time techniques integrate the amplification and analysis steps of the PCR reaction by monitoring the amount of DNA produced during each PCR cycle (Figure 22.1).1 The technology is still relatively novel, with the first practical real-time fluorescence-based quantitative PCR method, the 5′ nuclease assay, developed in 1996.2 Nevertheless, its conceptual simplicity,
sensitivity, specificity, and wide dynamic range have made real-time PCR-based assays the gold standard for the detection and quantification of both DNA\(^3\) and RNA.\(^4\) There are currently five main chemistries that use fluorescent dyes to monitor PCR reactions in real time during the PCR, although several others have been reported and new ones are being developed all the time (Table 22.1). Fluorescent reporting chemistries can be grouped into two types: those that are nonspecific and those that are specific. The former include intercalating dyes, e.g., SYBR-Green-I that bind to any double stranded (ds) DNA generated during the PCR reaction and emit enhanced fluorescence. These assays have two advantages: (1) they can be incorporated into optimized and long-established protocols that use legacy primers and experimental conditions, and (2) they are significantly cheaper as there is no probe-associated cost. Furthermore, they can yield quasi-template specific data if DNA melting curves are used to identify specific amplification products. Additional nonspecific chemistries are being developed and may be suitable for niche applications. Target-specific analysis

FIGURE 22.1 Typical amplification plots obtained using a real-time PCR assay. An amplification plot is the plot of fluorescence signal vs. cycle number. The fluorescence signal during the initial 12 to 14 cycles of the PCR reaction is below the instrument’s detection threshold and defines the baseline for the amplification plot. An increase in fluorescence above the baseline indicates the detection of accumulated PCR product, and the threshold cycle (\(C_T\)) is defined as the cycle number at which the fluorescence passes that threshold. A plot of the log of initial target copy number for a set of standards vs. \(C_T\) is a straight line. Quantification of the amount of target in unknown samples is accomplished by measuring the \(C_T\) and using the standard curve to determine the starting copy number, a process carried out by the software supplied with the real-time PCR detection instrument. The amplification plots for two replicates of two samples show the good reproducibility of the assay.

<table>
<thead>
<tr>
<th>TABLE 22.1</th>
<th>Current Fluorescence-Based Real-Time Chemistries</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonspecific Chemicals</strong></td>
<td><strong>Specific Chemicals</strong></td>
</tr>
<tr>
<td>Intercalating dyes</td>
<td>TaqMan™</td>
</tr>
<tr>
<td>AmpliFluor™</td>
<td>Hybridization</td>
</tr>
<tr>
<td>Quencher-labeled primers</td>
<td>Molecular Beacons</td>
</tr>
<tr>
<td>Lux™ primers</td>
<td>Scorpions™</td>
</tr>
<tr>
<td></td>
<td>Lanthanide</td>
</tr>
<tr>
<td></td>
<td>ResonSense™</td>
</tr>
<tr>
<td></td>
<td>Angler™</td>
</tr>
<tr>
<td></td>
<td>Hybeacons™</td>
</tr>
<tr>
<td></td>
<td>Cationic conjugated PNA</td>
</tr>
<tr>
<td></td>
<td>Light-up™</td>
</tr>
<tr>
<td></td>
<td>Eclipse™</td>
</tr>
<tr>
<td></td>
<td>Cyclicons™</td>
</tr>
</tbody>
</table>

\[ \Delta R_0 \]

\[ C_T \text{ sample 1} \]

\[ C_T \text{ sample 2} \]
requires the design and synthesis of one or more custom-made fluorescent probes for each PCR assay, and specific chemistries fall into two categories: (1) those that use nonstructured linear probes (e.g., TaqMan) and those that use structurally constrained constructs (e.g., Scorpions). All chemistries depend on the hybridization of one or two amplicon-specific hybridization probes and all are highly specific since nonspecific amplification due to mispriming or primer-dimer artefacts does not generate a signal and is ignored by the fluorescence detector. This obviates the need for post-PCR Southern blotting, sequence analysis, or melting curves to confirm the identity of the amplicon. Another advantage over intercalating dyes is that the probes can be labeled with different, distinguishable reporter dyes that allow the detection of amplification products from several distinct sequences in a single PCR reaction (multiplex).

The problem is that while quantification per se is simple, the interpretation and reporting of that quantitative data is not. In principle there are two ways to quantitate mRNA levels: either relative to some internal control or absolute per cell number, total RNA or unit mass of tissue.

II. TYPES OF QUANTITATIVE PCR ASSAYS

A. Relative Quantification

Relative quantification determines the changes in steady-state mRNA levels of a gene across multiple samples and expresses it relative to the levels of a coamplified internal control mRNA. During the RT-PCR assay, target C<sub>T</sub> are compared directly to reference C<sub>T</sub> and the results are expressed as ratios of the target-specific signal to the internal reference. This produces a corrected relative value for the target-specific mRNA product that can be compared between samples for an estimate of the relative expression of target mRNA in those samples. For example, c-myc mRNA levels relative to GAPDH mRNA levels are n times more or less in colorectal tumor cells compared with paired normal colonic biopsies (Figure 22.2A). An evaluation of amplification efficiency is essential as this has a major influence on the accuracy of any calculated expression result. Several models have been published that correct for efficiency using various algorithms and claim to allow a more reliable estimation of the real expression ratio. However, a crucial flaw with this approach remains in that the most common reference mRNAs are transcribed from so-called housekeeping genes whose expression is regulated and whose levels usually vary significantly with treatment or between individuals (see text below). Furthermore, if the relative levels of the housekeeping and target genes vary by orders of magnitude, the former may have entered its plateau phase by the time a C<sub>T</sub> for the target becomes apparent. This is likely to interfere with the accurate quantification of the target mRNA. However, relative quantification can generate useful and biologically relevant information when used appropriately. For example, an experiment designed to compare the degree of activated T-cell infiltration in colorectal cancers might usefully measure Interleukin-2 receptor (IL-2R) mRNA levels relative to those of a T-cell-specific marker (CD3 or CD8).

B. Absolute Quantification

This method is based on the use of an external standard dilution series with predetermined known concentrations, and results are expressed as actual numbers of mRNA or DNA molecules. Its most obvious application is in quantifying tumor cells or infectious particles such as viruses or bacteria in body fluids, but in the absence of suitable housekeeping genes that can act as normalizers, it is also usefully applied to quantitate changes in mRNA levels. The accuracy of absolute quantification depends entirely on the accuracy of the standards. In general, standard curves are highly reproducible and allow the generation of specific and reproducible results. Nevertheless, it is difficult to calibrate these standards so that they permit universal, absolute quantification, and results may not be comparable with those obtained using different probe/primer sets for the same markers and most certainly will be different from results obtained using different techniques. Furthermore, external.
standards cannot detect or compensate for inhibitors that may be present in the samples. For this it is necessary to spike the sample with an internal control. Ideally, this will be a synthetic amplicon, either amplified in a separate reaction or coamplified with the target using a different fluorophore to detect its amplification product.

Standard curves can be constructed using in vitro T7-transcribed sense-RNA transcripts or single-stranded sense-strand oligodeoxyribonucleotides. The standard curve is generated by performing serial dilutions of the standard and assaying each dilution together with positive and negative control reactions. To maximize accuracy, the dilutions are made over the range of copy numbers that include the amount of target mRNA expected in the experimental RNA samples. The $C_T$ value is inversely proportional to the log of the initial copy number. Therefore, a standard curve is generated by plotting the $C_T$ values against the logarithm of the initial copy numbers. The copy numbers of experimental RNAs can be calculated after real-time amplification from the linear regression of that standard curve, with the $y$-intercept giving the sensitivity, the slope, and the amplification efficiency.

III. CRITICAL FACTORS

A. Normalization

Biologically meaningful reporting of target mRNA copy numbers requires accurate and relevant normalization to some standard. Clearly, the quality of quantitative data cannot be better than the standards.
Meaningful Quantification of mRNA Using Real-Time PCR

quality of the normalizer and any variation in the normalizer will obscure real changes and/or produce artificial ones. RT-PCR–specific errors in the quantification of mRNA transcripts are easily compounded by any variation in the amount of starting material between samples. This is especially relevant when the samples have been obtained from different individuals or when comparing samples from different tissues, and will result in the misinterpretation of the expression profiles of the target genes. Consequently, the question of appropriate standardization arises and constitutes one of the most critical aspects of experimental design.

The accepted method for minimizing these errors and correcting for sample-to-sample variation is to amplify, simultaneously with the target, a cellular RNA specified by a housekeeping gene such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH) that serves as an internal reference against which other RNA values can be normalized. However, it is now clear that there is no single RNA with a constant expression level among different tissues of an organism and that their use as internal calibrators is inappropriate. Ribosomal RNA (rRNA), which makes up the bulk of a total RNA sample, has been proposed as an alternative normalizer and its levels remain fairly constant during serum-stimulation of tissue culture cells. However, reservations remain concerning its expression levels, transcription by a different RNA polymerase, and possible imbalances in relative rRNA to mRNA content in different cell types.

For some applications, e.g., when preparing RNA from tissue culture cells, nucleated cells, or microdissected cells, it is a straightforward approach to normalize mRNA levels to copies per cell number. Alternatively, normalization to total cellular RNA appears to be relatively reliable since there is little variability in total cell RNA content within the same tissue type between individuals. However, there are obvious limitations to this approach as total RNA levels may be increased in highly proliferating cells and this will affect the accuracy of any comparison of absolute copy numbers between normal and tumor cells. Nevertheless, combining absolute quantification using a target-specific standard curve with reporting of copy numbers relative to total RNA is as close as it is possible to achieve precise and biologically meaningful quantification (Figure 22.2B) and allow straightforward comparison between different laboratories.

B. QUANTIFICATION AND QUALITY ASSESSMENT OF RNA

Normalization to total RNA content requires accurate quantification of the RNA, and the most commonly used method, absorbance measurement at OD\textsubscript{260} in a spectrophotometer, may not be sufficiently accurate for this purpose. In addition, it can waste significant amounts of precious RNA. The RiboGreen RNA quantification assay relies on a proprietary dye that exhibits significant fluorescence enhancement on binding to nucleic acids and can be detected in a spectrofluorometer, fluorescence microplate reader, or filter-based fluorometer. However, it does not provide qualitative information. Both methodologies are likely to be superseded by the recent introduction of the RNA 6000 LabChip®, which allows the integration of the quantification of RNA samples with a quality assessment in one rapid step. It is based on microfluidics technology, and a LabChip cassette consists of a series of interconnected microchannels and reservoirs embedded in a palm-sized chip architecture. Migrations through the channels are monitored and controlled using the benchtop Agilent 2100 Bioanalyzer Instrument. There are two types of LabChip, the standard LabChip for accurate quantification above 50 ng/µl of RNA and the Pico LabChip® for concentrations of RNA down to about 200 pg/µl. Both calculate the ratio of 28S:18S rRNA in the sample to provide a simultaneous qualitative assessment for each sample. Results can be viewed as gel-like images, electropherograms, or in tabular formats. The importance of obtaining qualitative information is illustrated by Figure 22.3, which shows significant differences in absolute c-myc mRNA copy numbers in good and poor RNA samples extracted from identical tissue samples.

An important point to consider when using total RNA for normalization is the lack of internal control for RT or PCR inhibitors such as those reviewed in Chapter 5 by Rådström et al. All quantitative methods assume that the RNA targets are reverse transcribed and subsequently amplified with similar
C. REPRODUCIBILITY

Optimization and consistency are as critical for obtaining reproducible results using real-time RT-PCR as they are for conventional methods. However, real-time RT-PCR assays are significantly less variable than any conventional RT-PCR protocol which is subject to significant error. Where measured, the coefficient of variation for C\textsubscript{T} data has been shown to be very low at less than 2% for the TaqMan\textsuperscript{2,20} and as low as 0.4% for the Lightcycler\textsuperscript{21} which is significantly better than the 14% reported for conventional RT-PCR.\textsuperscript{22} Reproducibility is influenced by parameters such as distribution statistics (Poisson’s law),\textsuperscript{23} and C\textsubscript{T} data are less reproducible when working with very low copy numbers due to the stochastic sampling effects.\textsuperscript{24} Particle distribution statistics predict that it will require a much higher number of replicates to differentiate 5 from 10 copies of RNA than to differentiate of 5,000 from 10,000 copies. Of course, this emphasizes the importance of repetitive testing in clinical samples, and one of the strengths of these assays is the ease with which it is possible to determine multiple C\textsubscript{T} values for every sample, which encourages replicate determinations of the same sample and permits the application of statistical analyses to the quantification procedure.

D. RELEVANCE

Normal biopsies contain a range of different cell types, a problem exacerbated in heterogeneous tumor samples that also include normal and inflammatory cells as well as diversely evolved cell populations. In addition, normal cells adjacent to a tumor may be phenotypically normal but genotypically abnormal or exhibit altered gene expression profiles due to their proximity to the tumor.\textsuperscript{25} Hence expression profiling of such biopsies provides a composite of the whole population and cannot identify expression limited to subpopulations or sections of the biopsy. This may result
in the masking or loss of the expression profile of a specific cell type, or it may be ascribed to and dismissed as illegitimate transcription because of the bulk of the surrounding cells. Nevertheless, most \textit{ex vivo} RNA extractions and subsequent analyses are carried out from such biopsies with little regard for the different cell types contained within that sample. The critical importance of microdissection for maximizing the accuracy of quantitative gene profiling of individual cells or cell populations is demonstrated by the significant differences that have been detected in the gene expression profiles of microdissected and bulk tissue samples.

Laser capture microdissection (LCM) is the most powerful technique for extracting pure subpopulations of cells from heterogeneous \textit{in vivo} cell samples for detailed molecular analysis and is described in more detail in Chapter 4 by Rose G. Mage et al. Quantitative isolation of RNA from such small samples is possible, and mRNA can be isolated and accurate and reproducible expression levels quantified from archival paraffin-embedded tissue specimens even after immunohistochemical staining. Therefore, it may now be possible to use a single sample for immunocytochemistry, \textit{in situ} hybridization, as well as quantitative RT-PCR or microarray analysis.

\section*{E. \textbf{R}\textit{E}\textbf{V}ERSE \textbf{T}RANSCRIPTION}

The RT step is the least monitored yet crucially important part of any protocol design to generate sensitive and accurate quantification of steady-state mRNA levels. First, the amount of cDNA produced by the reverse transcriptase must accurately represent RNA input amounts. Therefore, the dynamic range, sensitivity, and specificity of the enzyme are prime considerations for a successful RT-PCR assay. Second, RT reactions are usually carried out between 40 and 50°C. These relatively low temperatures can result in nonspecific priming by both forward and reverse primers, which is a particular problem with very low concentrations of starting template. Here, such nonspecific side reactions can outcompete the amplification of the desired product and, if the concentration of genuine target is sufficiently low, inhibit it completely. Third, G/C- or secondary structure-rich mRNA templates also pose problems for reverse transcriptases transcribing at 40 to 50°C. This is because such templates can cause the enzyme to stop, dissociate from the RNA template, or skip over looped-out regions of RNA. Using a thermostable polymerase such as \textit{Tth} or AMV-RT at elevated temperatures reduces the secondary structure of the RNA template and can help with the read-through of the polymerase. Other strategies for dealing with RNA structure include supplementing the reverse transcription reaction with additives such as dimethyl sulphate (DMSO), sodium pyrophosphate, betaine, or trehalose which disrupt secondary structure, stabilize the enzyme, or prevent premature termination. However, these additives behave differently with different enzymes, templates, and reaction conditions, and their use cannot be recommended globally.

\section*{IV. CONCLUSIONS}

The comparative ease with which real-time RT-PCR assays can generate quantitative data has created the impression that this technique is simple, speedy, sensitive, and specific and that data can be subjected to objective statistical analysis. Nevertheless, significant doubts remain about the relevance and comparability of real-time RT-PCR data, and statistical analyses of the numerical data may obscure the actual results, allowing misinterpretation. The uncritical use of housekeeping genes to report relative changes of mRNA levels must be discouraged. Today’s main challenge is to develop experimental protocols and designs that are rigorously controlled and which allow meaningful global comparisons.

\section*{ACKNOWLEDGMENTS}

I would like to thank Becki Hands for carrying out the experiment in Figure 22.3 and acknowledge support from Bowel and Cancer Research.
REFERENCES

Meaningful Quantification of mRNA Using Real-Time PCR

Absolute Quantification of Specific Nucleic Acids by RT-PCR Using a Nonlinear Mathematical Model for Data Analysis

Huong L. Vu, Serge Troubetzkoy, Vu Q. Nguyen, Michael W. Russell, and Jiri Mestecky

CONTENTS

I. Introduction ................................................................. 235
II. Example ........................................................................ 236
   A. Experimental Problem .............................................. 236
   B. Generation of RNA Mimic and Reference Target ................. 237
      1. Primer Design .................................................... 237
      2. RT-PCR ........................................................... 237
      3. Ligation-PCR ..................................................... 238
      4. *In Vitro* Transcription ......................................... 238
   C. Competitive RT-PCR ................................................. 238
   D. Data Analysis with the PCRFIT Program ......................... 239
      1. Input Data ......................................................... 239
      2. Output Data ...................................................... 239
      3. Assessment of Assay Error .................................... 240
      4. Determination of Copy Number ............................... 242
III. Discussion ................................................................. 242
Acknowledgments .......................................................... 243
References ........................................................................ 243

I. INTRODUCTION

Quantification of specific nucleic acids plays a key role in the study of gene expression and in the clinical analysis of viral loads such as for human immunodeficiency virus (HIV-1) and hepatitis C virus (HCV). Among methods developed for this purpose, competitive (RT)-PCR, which involves reverse transcription (RT) of RNA to cDNA followed by polymerase chain reaction (PCR), is the most popular for its high sensitivity and relatively low cost. However, in the past decade the method has not improved much to overcome its main drawbacks: the difficulty of generating valid internal standards and the controversy over its accuracy.
Due to complex kinetics, competitive RT-PCR relies on a coamplified RNA/DNA internal standard (competitor, mimic) to correct for reaction efficiency. An ideal mimic should have the same kinetics as the target during the RT-PCR reaction and should be distinguishable from the target PCR product during the detection step, usually by amplicon size (homologous mimic) or by sequence (nonhomologous mimic). It is easy to see that the two requirements are contradictory and cannot be met directly in a single mimic sequence. However, these concerns can be addressed to a large extent by a new approach using two synthesized sequences as follows. A mimic with an internal deletion can be easily separated from the native amplicon by gel electrophoresis, while a reference target identical to the native amplicon can simulate exactly the kinetic behavior of the natural target. These two sequences, RNA/DNA mimic and reference target, can be simply generated at the same time by RT-PCR followed by ligation and/or in vitro transcription. No cloning, restriction analysis, or introduction of nonhomologous sequences is required. For absolute quantification of nucleic acids, two parallel competitive RT-PCR assays are set up where the mimic is coamplified with the reference target in one set and with the natural target in the other. The former enables us to determine the difference between reference target and mimic amplification efficiencies since their initial amounts are known in advance (measured by spectrophotometry). The difference is subsequently used for correcting the results obtained from the latter, i.e., quantitative assays of the natural target.

A PCR tube, from a kinetics point of view, represents a dynamical system, i.e., a system whose parameters such as composition, temperature, pressure, and volume are functions of time. The main purpose of studying a dynamical system is to predict values of certain parameters at a certain time point based on measured values available at another time point. In the case of quantitative PCR, the task is simply to predict the input (initial) concentration of an amplified sequence based on its output (final) concentration. Whether this backward prediction is feasible depends on the nature of PCR (its predictability) and on our ability to understand it. PCR has been considered to be kinetically difficult to control, and thus would be unsuitable for quantitative applications. However, our studies (unpublished), using dynamical system theory, have shown that PCR is intrinsically predictable and can, therefore, serve as a quantitative method. The second question concerns how the output concentration of an amplified sequence depends on the input concentration and on time, i.e., on the number of PCR cycles. The task is commonly simplified by setting one of the two independent variables to constant. For example, in a real-time PCR tube the input concentration is fixed and the output concentration is monitored as a function of time. By contrast, in end-point competitive RT-PCR, the input concentration changes in a dilution series while the reaction time is fixed as the number of cycles. In theory, PCR with constant amplification efficiency would have exponential kinetics which can be linearized through logarithmic transformation. In reality, due to a saturation effect, PCR amplification efficiency varies systematically from cycle to cycle even assuming the absence of random error. Therefore, a nonideal or “normal” PCR follows more complex polynomial kinetics which is not linearizable. There are two solutions to this problem: either to restrict the product detection step within the exponential phase to ensure linearity, as is done in real-time PCR, or to establish a nonlinear mathematical model describing PCR kinetics which would include the saturated phase, as reported previously. We have implemented the new mathematical model into a computer routine for a nonlinear Least Squares Fit method and created a self-executable program, PCRFIT, for practical data analysis. The new approach extends the validity of quantitative competitive RT-PCR into the saturated phase without requiring a new PCR instrument and additional costly reagents such as labeled probes.

II. EXAMPLE

A. EXPERIMENTAL PROBLEM

The most frequently encountered task of quantitative RT-PCR is to determine the number of mRNA transcripts in a given sample. In this example, we analyze the number of human J-chain mRNA transcripts present in 500 ng of total RNA isolated from the DAKIKI cell line.
Absolute Quantification of Specific Nucleic Acids by RT-PCR

**B. Generation of RNA Mimic and Reference Target**

1. Primer Design

A set of four primers, P1 to P4, were selected from the sequence of human J-chain cDNA as shown in Figure 23.1. The distance between P1 and P4 determines the size of the target amplicon which is preferably in the range of 400 to 600 basepairs (bp) for optimal in vitro transcription and gel separation. The distance between P2 and P3 decides the length of the middle fragment M which will be deleted from the sequence of the corresponding mimic. Within this size range and deletions of 70 to 100 bp, the mimic can be easily discerned from target by agarose gel electrophoresis. The primers (Table 23.1) are located in more than one exon. They have appropriate internal stability, i.e., higher stability at 5' end than at 3' end, as recommended by the Oligo 4.0 program (National Biosciences, Plymouth, MN). Primer T7-P1 has the promoter sequence of T7 RNA polymerase attached at the 5' end (lower case letters). Specificity of the primer sequences was checked using the BLAST program (NCBI, Bethesda, MD).

2. RT-PCR

Reverse transcription was performed with approximately 500 ng of DAKIKI total RNA using oligo dT as a primer, followed by three PCR reactions with the corresponding primer pairs T7-P1 /P2, P3/P4, and T7-P1 /P4 (Figure 23.1). Thus, three DNA fragments were obtained: U(Upstream, 173 bp plus 20 bp), D (Downstream, 298 bp), and UMD (Upstream + Middle fragment + Downstream, 560 bp plus 20 bp) as the T7 promoter sequence from the T7-P1 primer added 20 bp to the corresponding PCR products.

![FIGURE 23.1 Generation of RNA internal standard and reference target.](image)

**TABLE 23.1 J-Chain Primer Sequences**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Nucleotide Sequence</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>5'-GCCCAAGAAGATGAAAGGATTG</td>
<td>Exon 2</td>
</tr>
<tr>
<td>T7-P1</td>
<td>5'-taatacgactcactatagggGCCCAAGAAGATGAAAGGATTG</td>
<td>Exon 2</td>
</tr>
<tr>
<td>P2</td>
<td>5'-ATTGTTGAGGTGGGATCAGAGA</td>
<td>Exon 3</td>
</tr>
<tr>
<td>P3</td>
<td>5'-CCCAGAGCAATATCTGTGATGA</td>
<td>Exon 4</td>
</tr>
<tr>
<td>P4</td>
<td>5'-GGTGCCAGGGAGTTGTTTAC</td>
<td>Postexon 4</td>
</tr>
</tbody>
</table>

www.taq.ir
3. Ligation-PCR

The U and D fragments were phosphorylated by forward reaction with T4 polynucleotide kinase
and ligated together with T4 DNA ligase (Promega Inc., Madison, WI). The ligated DNA was then
amplified by 35 PCR cycles with primer pair T7-P1 and P4 producing fragment UD (471 bp plus
20 bp). This new PCR product has an internal deletion of 89 bp (fragment M) in comparison with
the amplicon from the natural target and can serve as a DNA template for \textit{in vitro} transcription to
make an RNA mimic or directly as a DNA mimic (Figure 23.1).

4. \textit{In Vitro} Transcription

RNA mimic and reference target were synthesized from the corresponding DNA templates, the
amplified fragments UD and UMD, respectively (Figure 23.1), by \textit{in vitro} transcription with T7
RNA polymerase. The transcripts were then digested with DNase I and subsequently purified on
RNeasy columns (Qiagen, Chatsworth, CA). The purified transcripts were quantified by measuring
OD$_{260}$, aliquoted, and stored at $-70^\circ$C. A solution of yeast RNA (Ambion, Austin, TX) in nuclease
free water (20 ng/µl) was used for dilution of the transcripts. Working solutions of transcripts (5
µl aliquots of 1 ng/µl concentration) were prepared and stored at $-70^\circ$C for single use.

For quantitative purpose it is necessary to determine the number of molecules in a given mass
unit, e.g., the number of RNA molecules in one nanogram of the synthesized transcripts. First,
relative molecular mass (M$_r$) of the single stranded (ss) RNA is calculated using the average M$_r$
of nucleotides$^{20,21}$

$$
M_r \text{ of ss RNA} = \text{number of nucleotides} \times 340 \text{ Da}
$$

For example:

$$
M_r \text{ of J-chain mimic} = 471 \times 340 \text{ Da} = 160,140 \text{ Da} = 160.14 \text{ kDa}
$$

Second, the mass unit is converted into the copy number of transcripts as follows (in the case of
J-chain mimic):

$$
1 \text{ mol} = 6.0221 \times 10^{23} \text{ copies} \equiv 160.14 \times 10^3 \text{ g}
$$
$$
1 \text{ g} \equiv 6.0221 \times 10^{23} \text{ copies/160.14} \times 10^3 = 3.76 \times 10^{18} \text{ copies}
$$
$$
1 \text{ ng} \equiv 3.76 \times 10^9 \text{ copies}
$$

Thus, in this case, 1 µl of the working solution (1 ng) contains $3.76 \times 10^9$ copies of the transcript.

C. Competitive RT-PCR

Due to the highly sensitive nature of RT-PCR, each experiment needs to be performed in replicate
to reveal the reliability of the method and the reproducibility of the result. For each quantitative
competitive RT-PCR assay we prefer to set up two different dilution series where the theoretical
ratio between the target RNA amounts in the two series equals 2, rather than to duplicate one
dilution series. In this way, we actually simulate a twofold change in mRNA level which is usually
considered as the threshold for biologically significant changes. Furthermore, two different dilution
series cover a wider range than two identical dilution series.

Figure 23.2 shows an example of competitive RT-PCR for 500 ng and 250 ng of DAKIKI total
RNA as series A and B, respectively. A similar experiment was performed for $2 \times 10^6$ and $1 \times 10^6$
copies of reference target RNA in C and D series, respectively (data not shown). In this reverse
transcription the lower primer P4 was used instead of oligo dT since the mimic RNA does not have
a poly-A tail. Since the following data analysis step requires at least five good data points, it is
Absolute Quantification of Specific Nucleic Acids by RT–PCR

239

recommended to set up at least seven RT-PCR reactions in each dilution series, allowing for experimental errors. In order to minimize tube-to-tube variation and pipetting labor, the target RNA was added to the RT master mixture and mixed well before being distributed into individual tubes in the dilution series. The denaturing time was kept minimal in order to avoid DNA damage during 35 PCR cycles. For the best results gel electrophoresis was run in the same day, avoiding random effects of storage condition on PCR products. The gel was scanned on a digital imaging system (AlphaImager 2000; Alpha Innotech) by the 1D-Multi (Line Densitometry) method. While making exposure adjustments, the color saturation was checked to ensure that the image was not oversaturated or underexposed.

D. DATA ANALYSIS WITH THE PCRFIT PROGRAM

1. Input Data

An input data file for PCRFIT can be created as a text file simply by using the NotePad program or PCRFIT’s menu under Windows 95 (or higher). Each data point (see the Data Example file in program Supplement) is given by the starting number of mimic copies (the X value) and the ratio between the scanned densities of target and mimic bands (the Y value). When the data file is open in the program, the corresponding graph automatically appears on the screen. The Edit/Options menu enables users to adjust the parameters, e.g., the lengths of the amplified target and mimic fragments, and the scale of the X-axis.

2. Output Data

The main result provided by PCRFIT is the equimolar point EP (X_{EP}, Y_{EP}) (Figure 23.3A and Figure 23.3B) where Y_{EP} is, in our case, equal to the ratio between LT (the length of target amplified fragment) and LM (the length of mimic), and X_{EP} is the X-value corresponding to Y_{EP} on the graph. In general, the density ratio at the equimolar point (Y_{EP}) can be expressed as follows:

\[ Y_{EP} = \frac{T_{EP} \times DT}{M_{EP} \times DM} \]

where T_{EP} and M_{EP} are the numbers of target and mimic molecules, respectively. DT and DM are the fluorescence densities of a target and a mimic molecule, respectively. Since T_{EP} = M_{EP}, it follows that Y_{EP} = DT/DM. If PCR product is detected by an intercalating dye such as ethidium bromide

\[ \text{FIGURE 23.2} \text{ Quantitative competitive RT-PCR for human J-chain mRNA with primer pair P1/P4. PCR products were detected by 1.5\% agarose gel electrophoresis and ethidium bromide staining. Lane 1: DNA marker. Lane 2 contains 500 ng of DAKIKI total RNA alone (positive control for natural target and negative control for mimic). Lanes 3 to 9: dilution series A containing constant 500 ng of DAKIKI total RNA with the mimic in 1.5-fold dilutions from } 25.0 \times 10^6 \text{ to } 2.2 \times 10^6 \text{ molecules. Lanes 10 to 16: dilution series B containing constant 250 ng of DAKIKI total RNA with the mimic in 1.5-fold dilutions from } 11.1 \times 10^6 \text{ to } 0.98 \times 10^6 \text{ molecules. Lane 17 contains 200 ng of K562 total RNA and } 0.98 \times 10^6 \text{ mimic molecules (negative control for natural target and positive control for mimic).} \]
or SYBR Green, then \( Y_{EP} = \frac{\text{constant} \times LT}{\text{constant} \times LM} = \frac{LT}{LM}. \) If PCR product are detected by labeled primers or hybridization probes then \( Y_{EP} = 1. \) This assumption is still valid when heteroduplexes are present: suppose that among \( T_n \) target molecules there are \( H_n \) of them involved in heteroduplex formation with an equal number of mimic molecules. Suppose further that heteroduplexes can be discerned from homoduplexes, e.g., with an Agilent 2100 Bioanalyzer (Lab-on-a-Chip Technology, Caliper Technologies, and Agilent Technologies), then the detected signal ratio is

\[
Y = \frac{(T_n - H_n) \times DT}{(M_n - H_n) \times DM}
\]

With \( T_{EP} = M_{EP} \) at the equimolar point, it follows again:

\[
Y_{EP} = \frac{DT}{DM} = \frac{LT}{LM}
\]

While \( X_{EP} \) is the equimolar X-value detected after RT-PCR (output or apparent EP), it is the input EP before RT-PCR that we are actually interested in. Due to a difference in target and mimic amplification efficiency the two equimolar points are not identical. This shift will be compensated for using the reference target as shown below. The consistency between data points in the dilution series is assessed by the chi-squared value shown next to the equimolar point. Representing the error of data fitting by the nonlinear mathematical model, the chi-squared parameters reflect the error of data fitting by the nonlinear mathematical model.

![Graph](image.png)

**FIGURE 23.3** Fitting competitive RT-PCR data by the program PCRFIT. Figure 23.3A and Figure 23.3B correspond to series A and B from Figure 23.2, respectively. The ratio between the scanned densities of target and mimic bands is plotted vs. the starting number of mimic copies. Symbols: empty circles are data points, the filled circle denotes the equimolar point \( EP \), a square is an outlier eliminated from the input data. Parameters: \( LT \) and \( LM \) are the lengths of target and mimic amplified fragments, respectively. Chi-squared parameters reflect the error of data fitting by the nonlinear mathematical model.

Results from two competitive RT-PCR experiments for natural target RNA (series A and B) and for reference target RNA (series C and D) are summarized in Figure 23.4. The relative errors of the assays (expressed in percentage) were calculated from the formula:
\[
\text{ERR\%} = \left( \frac{|R - R_{th}|}{R_{th}} \right) \times 100
\]

where R is the experimental ratio between the initial amounts of target RNA in the two series determined as the ratio of the corresponding X_{EP} values. R_{th} is the theoretical ratio which is equal to 2 in this case. For example, the relative error of the first experiment before and after elimination of the outliers in series A (Figure 23.3A) and B (Figure 23.3B) is

\[
\text{ERR\%} = \left( \frac{|9.934649/4.208914 - 2|}{2} \right) \times 100 = 18
\]

and

\[
\text{ERR\%} = \left( \frac{|9.145708/4.592947 - 2|}{2} \right) \times 100 = 0.4
\]

respectively. Similarly, the relative error of the second experiment decreased from 6 to 0.8% after elimination of an outlier in series C (Figure 23.4). Note, that elimination of the outliers results in better data fitting (chi-squared value drops) and at the same time in higher accuracy (relative error decreases). This indicates that the nonlinear mathematical model is adequate. When the same data sets are processed by linear regression, elimination of the same outliers led to inconclusive results. For the first experiment, the relative error increased from 5 to 15% (lower accuracy) while the values of coefficient r increased (better data fitting) from 0.981 to 0.996 (series A) and from 0.996 to 0.999 (series B). For the second experiment, after eliminating one outlier in series C, the relative error increased from 18 to 22% while the value of coefficient r negligibly decreased from 0.986 to 0.985. This fact indicates that coefficient r, the criterion for data fitting by the linear regression, does not reflect unambiguously the relative error of the quantitative assays. In other words, the linear model is not adequate for quantitative competitive RT-PCR.

\[\text{FIGURE 23.4} \quad \text{Results of the two described competitive RT-PCR assays evaluated by means of the nonlinear mathematical model. The values obtained before and after elimination of the outliers are framed by the dashed and the solid lines, respectively.}\]
4. Determination of Copy Number

So far, the data analysis step is common for both relative and absolute quantification. In the case of relative quantification the fold change of a specific mRNA level between given samples is determined as the ratio of the corresponding \( X_{EP} \) values. No correction is necessary for this type of quantification. For absolute quantification it is necessary to proceed with a further correction for the difference in amplification efficiency between the coamplified sequences. According to Figure 23.4 (series C and D) starting amounts \( 1 \times 10^6 \) copies of RNA reference target and \( 0.71 \times 10^6 \) copies of RNA mimic have been detected by competitive RT-PCR as equivalent quantities. This means that, under given experimental condition, the mimic sequence was more efficiently amplified than the target sequence. Without correction, the result of absolute quantification would be subject to a relative error of \( \left[ \frac{(1 \times 0.71)}{1} \right] \times 100 = 29\% \). Therefore, the \( X_{EP} \) value of series A and B should be corrected as \( 9.15 \times 10^6/0.71 = 12.89 \times 10^6 \) copies, and \( 4.59 \times 10^6/0.71 = 6.46 \times 10^6 \), respectively.

In the example shown, an antisense gene specific primer was used for reverse transcription. Thus, the cDNA derived from the reference and the natural (native) target RNA should have the same 5’ end. If PCR product is detected as a specific band then there should be minimal concern about the error caused by the different in 3’ ends of the two kinds of cDNA. Alternatively, a perfect reference target, i.e., a full-length transcript, can be generated by \textit{in vitro} transcription of a vector with full-length cDNA insert. For DNA quantification genomic DNA is a perfect reference target.

III. DISCUSSION

We have successfully applied the described method for absolute quantification of different human mRNA species such as interleukin-1\( \beta \), interleukin-6, immunoglobulin \( \alpha \)-chain, and J-chain in different cell lines. Except for the primer design, the procedure does not require any particularly specialized molecular expertise. The generation of RNA mimics can easily be accomplished within two working days and the transcripts, when properly stored and handled, are stable for years. The total cost for a PCR instrument, materials, and reagents is much lower in comparison with a real-time PCR system.

It has been assumed that, for an accurate quantitative RT-PCR assay, the data acquisition step should be restricted to the exponential kinetics phase. However, our study has shown that, using a new nonlinear mathematical model for data analysis, competitive RT-PCR can achieve excellent accuracy (relative error \( \leq 1\% \)) even in the saturated phase. The relation between the equimolar point and the fixed point of the function describing PCR kinetics\(^{13} \) suggests that the above mathematical interpretation is highly accurate. An RT-PCR assay without an internal control provides no fixed point and thus is unable to reach such accuracy even with real-time detection as discussed below. Introducing a mimic in real-time PCR would create a system of multiplex reactions with multiple labeled probes, requiring a multicolor fluorescence detection device where complexity and cost may outweigh the advantage of the method.

In comparison with real-time PCR, end-point PCR has a great advantage in regard to data acquisition because the output signal is detected at its peak where the signal-to-noise ratio is optimal. By contrast, the most common analysis method for real-time PCR uses the output signal at the detection threshold (e.g., where the signal-to-noise ratio is the lowest acceptable). For a given sample, the threshold cycle is usually determined as the intercept (crossing point) between a fluorescence baseline (a noise band, chosen by users) and the fluorescent signal curve. Since the fluorescence background level varies from tube to tube, setting a common baseline is sometimes difficult. In order to overcome this obstacle another data acquisition mode has been introduced,\(^{22} \) where threshold cycles are determined as “the first turning points” according to the values of the derivatives of the fluorescence curves. This improvement, however, does not resolve the intrinsic problem of low signal-to-noise ratio. Furthermore, since the fluorescence signal is not detected...
continuously, the derivatives of the fluorescence curves are technically undefined, and their approximated values are subject to certain error. Therefore, the calculated values of amplification efficiency were sometimes greater than 100%.23 The limitation of the data acquisition technique presents a “bottleneck” on the way toward an accurate mathematical model for real-time PCR. This is the reason why the model of a perfect PCR (100% amplification efficiency) is still in use.24

Perhaps, the main disadvantage of competitive RT-PCR is the time-consuming pipetting process during preparation of the dilution series. Recently, the rapid development of biorobotics technology for liquid handling, nucleic acid preparation (Qiagen, Beckman-Coulter, MWG Biotech, etc.), and agarose electrophoresis (E-gel 96, Invitrogen) opens up the possibility for high-throughput and fully automated competitive RT-PCR. Our computer program PCRFIT for data analysis would be especially useful for such an automated PCR system.

ACKNOWLEDGMENTS

We thank Dr. Satoshi Nishimura for collaboration during experiment performance, Dr. Jerry McGhee for kindly providing us access to a real-time PCR instrument, the LightCycler, Dr. Rudi Weikard for helpful discussion regarding mathematical modeling for real-time PCR, Dr. Frederik van Ginkel for critical reading the manuscript, and Lokeshbabu M. Thakkalapatti for assistance in preparing the manuscript.

REFERENCES

24 Single-Molecule PCR – Basic Protocols and Applications

Paul H. Dear

CONTENTS

I. Introduction................................................................................................................. 245
II. Basic Tool Kit for Single-Molecule PCR ................................................................. 246
   A. Expressing Amounts and Concentrations of DNA............................................... 246
   B. Some Simple Statistics......................................................................................... 246
   C. Contamination Control....................................................................................... 247
   D. Preparation and Handling of Template DNA....................................................... 247
   E. Basic Single-Molecule Amplification by Nested Two-Phase PCR.................... 248
   F. Hemi-Nested Amplification.................................................................................. 250
   G. Highly Multiplex Single-Molecule PCR ............................................................. 250
   H. PEP and Other Nonspecific Amplification Methods............................................ 250
III. Applications and Theory .......................................................................................... 252
   A. Reference Reagents, Validation and Quantitation of Single-Molecule
      Amplification........................................................................................................ 252
      1. Preparation of a Reference Template............................................................... 252
      2. Precise Quantitation of Reference Template................................................... 253
   B. Determination of Copy Numbers........................................................................ 254
      1. Results Expected and Interpretation................................................................ 255
   C. Happy Mapping.................................................................................................... 255
IV. Discussion and Further Applications of Single-Molecule PCR................................. 256
References ..................................................................................................................... 257

I. INTRODUCTION

With the invention of the PCR\(^1\) in the late 1980s, molecular biologists were able to easily amplify a defined segment of DNA from a complex mixture for simple detection or for further analysis. Although Southern blotting, cloning, and other methods had been able to fulfill some of these roles for many years,\(^2\) PCR provided unparalleled speed, selectivity, and versatility. It quickly became apparent that it could achieve outstanding sensitivity too, amplifying chosen targets from minute samples of genomic DNA by virtually unlimited factors.\(^3\) This great sensitivity has made possible the fingerprinting of forensic samples containing only hundreds or tens of cells, the sequencing of the few surviving DNA molecules in ancient biological specimens, and the diagnosis of genetic abnormalities in biopsy samples or preimplantation embryos. Several of these applications are covered in other chapters in this book.

If these techniques are carried to their logical conclusion, we reach a point where we can amplify individual molecules and look at DNA as the cell does — one molecule at a time. We find a paradigm shift from the conventional analogue approach (“how strong is this hybridization...
signal?, “how bright is the band on this gel?”) to a digital one (“how many molecules are there?”; “is this molecule of DNA present or absent?”). This in turn enables a range of analytical approaches such as unambiguous copy-number determination, haplotyping, or accurate genomic mapping.

This chapter introduces basic techniques for amplifying, handling, and detecting single molecules of DNA and some simple applications of this technology. Single-molecule PCR is not difficult but it does require a systematic and careful approach. It also requires a new way of thinking. Analysis at this level of sensitivity has been dubbed “homeopathic PCR,” but it is precisely the opposite — DNA ceases to be something continuous which can be diluted indefinitely and becomes a collection of countable particles, each of which is either present or absent.

II. BASIC TOOL KIT FOR SINGLE-MOLECULE PCR

A. Expressing Amounts and Concentrations of DNA

It is imperative, in single-molecule analysis, to stop thinking of DNA in terms of nanograms, picograms/microliter, and similarly quantitative measures as they quickly cease to be relevant. The appropriate units are, instead, the number of copies of a target sequence or the number of copies per microliter of solution.

A megabase of double-stranded DNA weighs about one femtogram (fg, $10^{-15}$ g), and so a single (haploid) copy of a mammalian genome (around 3000 Mb) weighs about 3 picograms (pg, $3 \times 10^{-12}$ g). One nanogram of human DNA, then, contains about 300 copies of the genome (and hence 300 copies of any nonrepeated sequence). A solution of 1 ng/µl of human DNA, therefore, contains about 300 copies of the genome per microliter, or 300 G/µl (using “G” to indicate one genome’s worth of DNA of the species in question). The genome of E. coli, by contrast, is about 5 Mb and weighs about 5 fg, so a solution of 1 ng/µl of this DNA contains about 200,000 copies of the E. coli genome per microliter (200,000 G/µl). When adapting protocols from one species’ DNA to that of another, it is important to deal in equivalent numbers of copies rather than equivalent masses of DNA.

B. Some Simple Statistics

In the following sections, it is important to bear in mind that “one genome equivalent” of DNA does not mean one complete copy of the genome. Instead, it means a mass of DNA equal to the mass of the genome but sampled at random. It is therefore necessary to understand some fairly simple statistics when dealing with samples that contain only a few molecules. Clearly, taking a series of samples which on average contain one molecule of template each does not mean that each of them will contain exactly one molecule. More probably, some will contain two or three, while some will contain none.

The Poisson distribution states that, if an infinite number of samples contain an average of $Z$ template molecules each, then the proportion ($P_0$) of samples containing no template molecules is given by:

$$P_0 = e^{-z}$$

where $e$ is the base of natural logarithms, 2.718. So, if we take one hundred 1-µl samples from a solution containing one molecule of template per microliter, then $Z = 1$, and hence:

$$P_0 = e^{-1} = 0.37$$

In other words, we should expect (on average) 37% of such samples to be negative if scored for that template.
Equation (1) can be inverted to find the average DNA concentration given the proportion of negative samples:

\[ Z = -\ln(P_0) \] (2)

For instance, if we analyze one hundred 2 µl DNA samples and find that only 72 of them give a positive result, then the proportion of negatives, \( P_0 \), is 28/100 or 0.28 and so:

\[ Z = -\ln(0.28) = 1.27 \]

That is, the DNA which we sampled contains 1.27 template copies per 2-µl sample, or 0.63 copies per microliter. Readers well-versed in statistics will appreciate that the value of \( P_0 \) measured experimentally (0.28 in the above example) is only an estimate of the true value since the proportion of negative samples is itself subject to a sampling error. (For example, analyzing another set of 100 samples from the same DNA stock might have given 70 or 73 positives rather than 72.) To a reasonable approximation, adequate for most purposes, the number of negatives, \( N_0 \), has a standard deviation of \( \sqrt{N_0} \).

C. Contamination Control

Other chapters in this book, particularly the one by Anne C. Stone (Chapter 1) dealing with the amplification of ancient DNA, address the control of PCR contamination. However, I would like to briefly reiterate some of these points and address contamination control from a different perspective. The protocols used in this chapter should yield a very strong product from a single template molecule, and it follows that even a single contaminant molecule of the same template will completely obscure the true result. Prevention of contamination must therefore be absolute.

The rules for avoiding contamination are simple: all operations must be performed under clean conditions, up to the point where you close the PCR tube or seal the PCR plate to prevent further ingress of contaminants. A dedicated work area (if possible, a small dedicated room) should be set aside for “clean” work and a laminar flow cabinet within this area reserved for critical operations. As far as possible, nothing should be brought into this area from the general laboratory — be aware that your laboratory notebook is likely to be amongst the most heavily contaminated items! Reagents should be bought readymade where possible, to reduce in-house handling and stored in single-use aliquots. HPLC-grade bottled water should be used, preferably aliquotted and stored frozen.

It should not be necessary to decontaminate newly bought plasticware or reagents by UV exposure or autoclaving: most commercially produced items are free from contamination. Exceptions to this rule are enzymes (such as Taq polymerase) which may be contaminated with DNA from the bacterial host used for their production. Unfortunately, there is no simple means to decontaminate enzymes rigorously.

D. Preparation and Handling of Template DNA

There are two points to bear in mind when working with very low quantities of DNA. First, almost all surfaces (pipette tips, microplate wells) will bind DNA to a greater or lesser degree. Therefore, reliable dilution of DNA becomes problematic below a certain concentration. To ensure reproducibility from experiment to experiment, the same protocol needs be followed in detail: diluting 5 µl of template in 50 µl of water may yield different results from diluting 50 µl in 500 µl. Second, very dilute DNA samples are exquisitely sensitive to nucleases. Therefore, they should be stored frozen where possible (but avoid repeated freeze–thaw cycles by storing single-use aliquots) and protected by buffers containing small amounts of EDTA.

Clearly, very dilute templates are as susceptible to contamination as any other reagent, so the later dilution stages (beyond a few tens of G/µl) should be done under clean conditions.
E. Basic Single-Molecule Amplification by Nested Two-Phase PCR

Under ideal conditions, a conventional PCR continued for 50 to 60 cycles can yield a detectable product from a single-template molecule. However, such amplification is critically dependent upon the primer in question and often unreliable. Moreover, the products of such a single-phase amplification are themselves ideal templates for subsequent reactions and thus greatly increase the risk of contamination by “PCR carryover” if they find their way into the clean area where reactions are set up. Multiplexing single-phase, single-molecule PCRs is also virtually impossible.

A more robust approach is nested PCR (Figure 24.1). An initial amplification to modest levels is performed using one pair of primers. The products of this reaction are then diluted and a sample is reamplified using a second pair of primers which prime internally to the first pair. This approach consistently produces strong amplification from single-template molecules and is fairly tolerant of less-than-optimal primer design. Also, spurious products from mispriming are less likely since the two phases provide a double selection for the correct product. Moreover, the products of the second-phase PCR cannot act as templates for the first-phase primers, greatly reducing the risk of contamination of later reactions. The first phase is set up in the clean area, but the dilution and setting up of the second phase can be done on the open bench in the main lab.

Primer design is straightforward, and many available programs will produce suitable primers (although some ingenuity may be required with many of them which will not allow you to design nested sets directly). In our laboratory, our general guidelines for all primer design are a length of 19 to 22 bases, a calculated melting temperature ($T_m$) of 55 to 60°C, with two G/C bases at the 3' end and one at the 5' end as “clamps.” Internal amplimer length is typically 80 to 300 bp, with external primers as close as possible to the internal ones. Nevertheless, many primer sets will still work well despite not meeting all of these criteria, provided that the overall $T_m$ is correct. It is not normally necessary to purchase primers purified by HPLC or gel electrophoresis — these steps introduce more chances for contamination.

---

**FIGURE 24.1** Nested PCR format. An external pair of primers (solid arrows) is used for initial amplification of the target to modest levels in a phase-one PCR. A sample of this product is then reamplified using internal primers (dashed arrows) to high levels in a phase-two PCR. The internal amplimer produced after the second phase lacks priming sites for the external primers and hence cannot act as a contaminant in subsequent experiments.
First-phase PCRs, set up under clean conditions, work best with a “hotstart” enzyme system (i.e., one in which there is no polymerase activity until the reagents have gone through the first round of denaturing). However, other enzymes and corresponding buffer systems can be used.

**Phase-one PCR reaction**

- Template DNA (or water for negative controls)
- 1 X “Gold” PCR buffer (Perkin–Elmer)*
- 200 μM each dNTP (Pharmacia)*
- 2 mM MgCl₂*
- 0.1 μl (0.5 U) *Taq* “Gold” (Perkin–Elmer)
- 1 μM each forward and reverse external primers**
- Water (HPLC grade) to 10 μl total volume

Reactions can be overlaid with one drop of mineral oil or performed oil-free in a suitable thermocycler in plates or tubes. Cycling conditions are:

- Initial denaturation 93°C for 9 min
- then 25 cycles of 94°C for 20 sec
- 52°C for 30 sec
- 72°C for 60 sec***
- hold at 4°C

The high magnesium concentration (2 mM), low annealing temperature (52°C), and relatively large amount of polymerase result in a fairly low-stringency reaction. However, this increases the success rate in priming on small quantities of template, and the second-phase PCR should provide enough selectivity to eliminate spurious amplification products. After cycling, the first-phase products are diluted to 200 μl with water, and 5 μl of this dilution is used as template in the second-phase reaction. The dilution and the setting up of the second phase are carried out in the main lab area.

**Phase-two PCR reaction**

- 5 μl diluted phase-one product
- 1 X “Gold” PCR buffer (Perkin–Elmer)
- 200 μM each dNTP (Pharmacia)
- 1.5 mM MgCl₂
- 0.05 μl (0.25 U) *Taq* “Gold” (Perkin–Elmer)
- 1 μM each forward and reverse external primers
- Water (HPLC grade) to 10 μl total volume

Cycling conditions are:

- Initial denaturation 93°C for 9 min
- then 33 cycles of 94°C for 20 sec
- 55°C for 30 sec
- 72°C for 60 sec

---

* These components can conveniently be prepared as a premix and stored at –20°C in single-use aliquots.

** Primer concentrations as low as 0.1 μM are normally just as effective and reduce the risk of introducing contaminants along with the primers.

*** An extension of 2 min at 68°C is more effective for extremely A/T-rich DNA.
The magnesium concentration and annealing temperature may be adjusted to suit the primers in use, but once again the double-selection afforded by a two-phase PCR means that even fairly low-stringency conditions (high magnesium, low annealing temperature) will generally yield clean products. The number of cycles can be increased to 35 or 36 if products are weak, but this should not generally be necessary. Increasing the number of second-phase cycles increases the risk of contaminants (introduced while setting up the second phase reactions) giving a visible product.

Of the second-phase PCR products, 5 µl should be checked on an ethidium- (or SyBr Green-) stained gel. Band intensity should not be dependent upon the number of template molecules present at the outset: a single template should be enough to saturate the reactions and produce a strong band. Several negative controls (no template) should be run in each experiment, and should, of course, give no product. Strong bands in one or some of the controls imply sporadic contamination (probably with single molecules) in the phase-one reactions; bands in all controls (especially if faint) imply heavier contamination that was introduced when setting up the phase-two reactions.

**F. Hemi-Nested Amplification**

To reduce primer costs, one of the external primers can be eliminated from the nested PCR reaction and a single (“internal”) primer used in both phases. For example, an external-forward and a reverse primer can be used in phase-one reactions, and an internal-forward primer with the same reverse primer used in phase-two. Other reaction conditions are identical to those used for a fully nested reaction. Hemi-nested reactions are usually just as sensitive as fully nested PCRs, though in rare cases there may be some loss in specificity (through loss of the selection provided by the fourth primer).

**G. Highly Multiplex Single-Molecule PCR**

Multiplexing of conventional, single-phase PCRs is problematic. As the number of primer pairs in a reaction is increased, the opportunities for mispriming and for the formation of primer-dimers rise dramatically, leading to spurious or failed amplification. Moreover, the amplimers must be chosen to be sufficiently distinct in size to be resolvable on gels.

The first-phase reactions of nested (or hemi-nested) reactions, however, can be highly multiplexed with no adverse effects. This is possible because the level of amplification (typically 25 cycles) is not high enough for product–product interactions to be severe, and because any products of mispriming areweed out by the second-phase reactions. Second-phase reactions, however, are normally monoplex (i.e., not multiplexed).

It is therefore possible to analyze the same minute DNA sample for many sequences simultaneously by performing a single multiplexed first-phase PCR and then using aliquots of the diluted product for successive second-phase reactions for each sequence in turn. Up to 600 external primer sets can be multiplexed in this way (J. Pachebat, personal communication). Reaction conditions are the same as for normal nested (Section II.E above) or hemi-nested (Section II.F above) PCR. However, when multiplexing more than a few tens of external primer pairs in the first phase, the magnesium concentration should be raised to 3 or even 4 mM (since additional primers bind magnesium) and the concentration of each primer reduced to 0.2 µM. Additional cycling in phase one (29 rather than 25 cycles) also allows the products to be diluted further (up to 3 µl) to provide the number of 5-µl subfractions necessary for up to 600 separate second-phase reactions.

**H. PEP and Other Nonspecific Amplification Methods**

Multiplex PCR allows the amplification of up to several hundred different sequences from a single minute DNA sample. However, all of these target amplimers must be chosen at the outset and their external primers included in the initial amplification. There is no opportunity to return to the same sample later and amplify additional sequences.
A solution to this problem is to use a nonspecific amplification step initially, ideally amplifying all of the sequences present in the sample by a large factor. Subfractions of this product can then be taken repeatedly for further analysis. Chapter 34 by Wells and Bermudez deals with a range of protocols for “whole genome amplification” (WGA), all of which are successful to a greater or lesser degree. Broadly, our own experiences agree with those of Wells and Bermudez, but true single-molecule analysis provides an even more stringent test of WGA protocols. We find that DOP-PCR, Tagged-PCR, and other variants which rely on priming by partially degenerate primers (e.g., Reference 6 and our own unpublished data) are strongly biased at the single-molecule level, amplifying less than 50% (often <10%) of target sequences from single molecules consistently. We are evaluating a linker– adaptor protocol but, for the moment, primer-extension preamplification (PEP) appears to be the simplest and most robust WGA protocol for single-molecule analyses. It relies on priming with a completely degenerate 15-mer (“N₁₅”) at low temperature and is apparently unbiased. In our hands, it reproducibly amplifies more than 90% of all specific sequences from single molecules. Its drawback, as Wells and Bermudez note, is that the level of amplification (which we estimate at 50- to 200-fold) is very poor.

Despite this limitation, PEP is still sufficient for most single-molecule analysis. A sample can be amplified by PEP and the products split into 20 to 35 subfractions, each still containing all the sequences present in the original sample. Each of these subfractions can then be analyzed using multiplex nested or hemi-nested PCR for several hundred specific sequences, allowing up to 20,000 amplimers to be scored — adequate for most purposes.

Our PEP protocol, optimized for single-molecule amplification, differs slightly from that of Wells and Bermudez. In particular, we find a smaller reaction volume to work better and, when amplifying single copies of very small genomes (and hence less than 1 pg of total template), we see better results if reactions are supplemented with a few picograms of “carrier” DNA from a different species. This carrier DNA is amplified along with the target, but this does not normally matter if the specific primers to be used afterwards are species-specific. We also find that the quality of the N₁₅ primer varies from supplier to supplier (and possibly from batch to batch).

**PEP amplification**

- Template DNA (or water for negative controls)
- 1 X PCR buffer II (Perkin–Elmer)
- 200 μM each dNTP (Pharmacia)
- 2.5 mM MgCl₂
- 0.2μl (1 U) *Taq* polymerase (“AmpliTaq,” Perkin–Elmer)
- 10 μM N₁₅ primer
- ~3 pg “carrier” DNA*
- Water (HPLC grade) to 5 μl total volume
- Overlay with light mineral oil (Sigma)

Cycling conditions are identical to those specified for PEP by Wells and Bermudez. After cycling, products should be diluted to 150 μl with HPLC water, and 5μl subfractions aliquotted and stored for subsequent two-phase (nested or hemi-nested) amplification. The dilution of the PEP products (as well as the first phase of the subsequent PCRs) must be done under clean conditions, requiring that the PEP reactions be brought back into the clean area. This should be done with caution since the outside of the microplate (or tube) will probably be heavily contaminated through contact with the thermocycler.

* Carrier is needed only if the expected mass of template DNA is less than 1 pg; it should be genomic DNA of a species unrelated to that under analysis.
III. APPLICATIONS AND THEORY

This section takes the reader through a few basic applications of single-molecule PCR, placing the above protocols into context. When beginning single-molecule PCR, it is important to systematically establish a set of reagents and procedures which are known to work routinely, including a “reference system” which will form the basis of monitoring later more complex experiments. The first of the following applications covers this reference system.

A. REFERENCE REAGENTS, VALIDATION AND QUANTITATION OF SINGLE-MOLECULE AMPLIFICATION

It is essential to be able to determine whether single molecules of DNA are being detected efficiently, but this raises a problem: it is difficult to isolate a known, single DNA molecule on which to test amplification. A partial — statistical — solution is to try to detect a target sequence in DNA at progressively greater dilutions with multiple replicates of each dilution. This needs to be done until a dilution is found which gives a signal from some samples (presumably those which contain one or two template molecules) but not from others (which are presumed to lack a template molecule). Even then it is possible that the supposedly limiting dilution contains several template molecules per sample, of which only a minority are being detected.

The solution to this problem is to try to detect, simultaneously, not one but two target sequences, known to lie only a few kilobases apart. Because of their proximity on the DNA, any molecule that carries the first sequence should also carry the second (unless the DNA is very badly fragmented). Therefore, efficient detection of both “reference sequences” should show them to be either both present or both absent in any given minute sample of DNA.

The two target sequences (which are referred to here as “A” and “B”) can be selected from any region of known single-copy sequence* in the genome in question and should be separated from each other by 1 to 3 kb. Nested (or hemi-nested) primers should be designed for each target, following the guidelines in Section II.E.

1. Preparation of a Reference Template

Working in the main laboratory, prepare a dilution of genomic DNA (in 10 mM Tris pH 7.5, 1 mM EDTA) containing ~100 genomes per microliter (100 G/μl). For mammalian DNA (~3000 Mb, and hence with a genome mass of ~3 pg), this is equivalent to 300 pg/ul. Multiple 100-μl aliquots of this diluted DNA should be frozen and stored at −80°C.

In the clean area, prepare further serial dilutions in water from one of these aliquots, to give an estimated 5, 2, 1, 0.5, and 0.25 G/μl. From each of these dilutions, set up 16 phase-one PCR reactions (Section II.E), each containing 1 μl of the appropriate serial dilution and all four of the external primers. A further 16 reactions containing water in place of the dilute DNA serve as negative controls; this gives a total of 96 reactions, conveniently accommodated in a 96-well thermocycler plate. Carry out phase-one amplification, followed by dilution and separate phase-two reactions for markers A and B on all 96 samples, then score the results on gels.

a. Results Expected

All 16 samples containing the highest DNA concentration (1 μl at 5 G/μl, i.e., 5 genome equivalents per sample) should give a clear band of the expected size for each of the two markers. The next dilution (2 genomes per sample) should also give clear bands for each marker in all or most of the

* In most genomes, particularly eukaryotes, there is a risk that any target sequence chosen may be duplicated or multicopy. In the absence of other information, the nontranslated regions of genes are likely to be single copy. If in doubt, carry out the analysis using a second pair of reference markers (“C+D”) from a different region of the genome and compare the results obtained with the two sets of markers.
samples if the dilutions were accurate and if amplification is efficient. With greater dilution (1 genome per sample and less), fewer and fewer of the 16 replicates should contain the target sequence and hence more should be negative for both of the markers. All 16 negative controls for each marker should also, of course, be blank.

It is unlikely that the dilutions will be perfectly accurate, and so bands may start to disappear from the replicates at a higher or lower dilution than predicted. Nevertheless, one of the dilutions should give between 4 and 8 positives out of the 16 replicates for each of the two markers. Importantly, amongst these 16 samples, most of them should contain either both markers or neither. If many of these samples contain marker A but not B, or vice versa, then detection of one or both of these markers is inefficient, and this must be rectified before proceeding (see text below).

Based on these results, one should be able to determine a dilution which will give approximately 50% positives (for each of the two markers) when analyzed in this way. Take a fresh aliquot of the ~100 G/μl stock, dilute it in water by this amount, and dispense it into a large number of 100-μl single-use aliquots for storage at –80°C. These aliquots will serve as a standard or reference template for all subsequent experiments.

2. Precise Quantitation of Reference Template

The reference template produced in A should contain around 0.5 G/μl (since 1-μl samples of it gave positive results in about 50% of the samples for either of the two reference markers), but this estimate is only approximate. For a more precise quantitation, carry out the two-phase PCRs for markers A and B on a total of 88 1-μl samples from an aliquot of the reference template, plus 8 negative controls, and score the results for each marker in all 96 samples.

a. Results Expected

All negative controls should be blank for both markers. Of the remaining 88 aliquots, approximately half should be positive for marker A and about half should be positive for marker B. As before, the majority of the samples should be either positive for both markers (“AB”) or negative for both (“ab”). If the number of samples containing only one marker (“Ab” or “aB”) is more than about a third of the number containing both (AB), this indicates that detection of one or both markers is inefficient. In this case, the PCR conditions should be optimized (for example, by adjusting the magnesium concentration or annealing temperature) or, if this proves unsuccessful, new primers should be designed until results show that both markers are being efficiently detected. If detection is efficient, then the exact concentration of the reference DNA can be calculated based on the Poisson distribution (Section II.B). The proportion of negatives ($P_0$) is best calculated as the average of the results for makers A and B. In other words:

$$P_0 = 1 - \frac{(N_a + N_b)}{(2 \times N_s)}$$

where $N_a$ is the number of samples positive for marker A, $N_b$ the number positive for marker B, and $N_s$ the total number of samples (88). For example, if 39 of the 88 aliquots are positive for A, and 40 are positive for marker B, then:

$$P_0 = 1 - \frac{(39 + 40)}{(2 \times 88)} = 0.537$$

and so:

* Readers familiar with linkage analysis should calculate the lod score between the two markers (see Reference 8 and Reference 9). Provided that the proportion of positives for each marker is between about 25 and 75%, a lod score of 8 or more should be seen between the two markers.
Hence, in this example, the reference DNA contains 0.60 genome equivalents in each 1-μl aliquot (0.30 G/μl).

Figure 24.2 (gels A and B) illustrate the results of a similar experiment, though in this example the quantitation of the reference markers was carried out as part of a larger analysis in which multicopy markers were also identified (see Section III.B).

B. DETERMINATION OF COPY NUMBERS

Having verified the system for detecting single-template molecules in the chosen species, this approach can be used for the precise determination of relative copy numbers of different sequences within the genome. All that is needed is to score the number of positives for the sequence in question on a set of samples of the dilute reference DNA and compare this with the number of positives obtained for the single-copy reference markers A and B on the same set of samples.

Prepare a set of 88 samples, each containing 0.5 to 0.8 genome equivalents of the reference DNA (Section III.A), and a further 8 negative controls. Amplify these in a phase-one multiplex PCR using external primers for the two reference markers (A and B) and for the marker (X) whose copy number is to be evaluated.* Dilute the products to 200μl and perform phase-two PCRs on 5μl subfractions for each of the markers in turn, using the appropriate internal primers. Analyze the products by gel electrophoresis and score the results for each marker.

* The reactions can be multiplexed further to determine the copy number of many markers in parallel. For simultaneous copy-number determination of very large numbers of samples the PEP protocol can be used to preamplify the initial samples, and these products split into subfractions as described in Section II.H.
1. Results Expected and Interpretation

Results of a typical analysis are shown in Figure 24.2. All negative controls are blank. Of the 88 samples, 44 are positive for reference marker A, and 45 for reference marker B. In all but one case, the same samples are positive for both of these tightly linked, single-copy reference markers, indicating efficient detection. Based on these results, the mean concentration of the single-copy sequences in the samples is calculated as follows:

\[
P_{0(A,B)} = 1 - \left( \frac{N_a + N_b}{2N_s} \right)
\]

\[
= 1 - \left( \frac{44 + 45}{2 \times 88} \right)
\]

\[
= 0.494
\]

and so:

\[
Z_{(A,B)} = -\ln(0.494) = 0.70
\]

In other words, the reference markers are present at an average of 0.70 copies per sample. Other single-copy markers would be expected to give a similar proportion of positive samples. Marker C, however, is present in 63 of the 88 samples. The corresponding calculation for this marker is:

\[
P_{0(C)} = 1 - \left( \frac{N_c}{N_s} \right)
\]

\[
= 1 - \left( \frac{63}{88} \right)
\]

\[
= 0.284,
\]

and so:

\[
Z_{(C)} = -\ln(0.284) = 1.26
\]

So, marker C is present at an average of 1.26 copies per sample. Comparison with the single-copy reference markers (0.70 copies per sample) implies that marker C is a duplicated sequence.

Clearly, markers present in triplicate or higher numbers of copies will tend to give positives in almost all of the samples analyzed by the above protocol and quantitation becomes much less accurate as the proportion of positives approaches 1.0. This is seen for marker D (Figure 24.2), which is positive in 82 of the 88 samples. In this case, the calculation gives a mean of 2.69 copies per sample which, compared with 0.70 for the single-copy reference sequences, implies a copy number of about 4 (i.e., a quadruplicated sequence); however, this estimate is prone to error. In such cases, the experiment should be repeated with less reference DNA. For example, using 0.3 genome equivalents of the reference DNA per sample would result in single-copy reference markers A and B being scored positive in about 25% of the samples; a double-copy marker (present at an average of 0.6 copies per sample) will be present in about 45% of samples, and a four-copy marker (present at an average of 1.2 copies per sample) in about 70% of the samples.

C. Happy Mapping

The quantitation of the reference DNA in Section III.A and Section III.B entailed a rudimentary form of genome mapping. The two reference markers A and B were found to cosegregate (to
occur in the same dilute DNA samples) because they were adjacent in the genome and thus likely to be found together on the same DNA fragments. Had their initial genome positions not been known, the results would have implied that they lay closer together than the average size of the DNA fragments.

This principle underlies a general method for genome mapping, known as HAPPY mapping.\(^9,10\) It can be used to accurately determine the order and spacing of large numbers of markers (sequence-tagged sites or STSs) in any genome. A complete discussion of HAPPY mapping is beyond the scope of this chapter. Briefly, however, a panel of about 90 DNA samples is prepared, each sample containing approximately half a genome equivalent of randomly fragmented DNA. The mean size of the fragments is chosen by the experimenter and determined by the degree of mechanical shearing or irradiation used to break the DNA. Hundreds or thousands of markers are scored on this panel by using PEP to achieve an initial preamplification, followed by highly multiplexed nested or hemi-nested PCR for groups of up to 600 markers. The data for all markers are then analyzed using software essentially similar to that employed for genetic or radiation-hybrid mapping. Markers which are close to one another in the genome (compared to the mean size of the DNA fragments) cosegregate strongly, the degree of cosegregation reflecting their physical proximity on the genomic DNA.

This approach has been applied to a number of genomes,\(^11–14\) and has proven robust and reliable. The key to its effectiveness is the ability to analyze samples of the genome by single-molecule PCR rather than relying on cloning or other biological methods to propagate them for analysis.

**IV. DISCUSSION AND FURTHER APPLICATIONS OF SINGLE-MOLECULE PCR**

Single-molecule PCR is a powerful tool with many applications. Broadly, it allows researchers to quantify DNA molecules precisely by counting them and to determine whether two sequences are present on the same DNA fragment without the need for cloning or other biological steps. It also allows the detection and analysis of very rare molecules in complex mixtures or their recovery from minuscule samples. The applications outlined above are simple uses of this tool for determining the copy number of sequences and mapping their locations on genomic DNA.

There are many other applications for such single-molecule analysis — some tested, some not. For example, the amplification methods described are applicable to the detection and analysis of very small numbers of parasites in clinical specimens\(^15\) or any other circumstance where sparse templates must be identified. We have not yet tried the use of these methods for the precise quantitation of rare expressed sequences, but there is no reason why this should not be possible following reverse transcription of mRNA.

Single-molecule approaches also make possible the determination of molecular haplotypes (that is, identifying which alleles of several linked loci are present on each of the two chromosomes of a diploid cell). For example, the alleles of two or more markers present on a single fragment of genomic DNA (isolated by limiting dilution) can be scored, to produce a haplotype which extends over the length of the fragment (for example, see Reference 16). More extended haplotypes can be built up using a variant of the HAPPY mapping approach. If marker alleles (rather than simply marker presence/absence) are scored on a HAPPY panel, then linkages are seen only between those alleles which lie on the same chromosome, and the entire map resolves into two haplotypes. These haplotypes can extend over chromosomal distances and across many markers, provided that the distance from each polymorphic marker to the next is not greater than the range of the mapping panel.
REFERENCES

I. INTRODUCTION

For those of us who study the expression of genes — the expressed genome or transcriptome — an array of state-of-the-art technologies exists that allows us to assess the expression of tens of thousands of genes simultaneously. These technologies, collectively called differential gene expression (DGE) technologies, allow scientists for the first time to investigate how global changes in gene expression contribute to a host of biological phenomena including cellular differentiation, development, aging, and progression of disease and injury. This chapter is intended to describe current technologies and assist in the selection of the appropriate currently available tools. The pivotal factors to be considered when choosing a method of gene expression analysis include, but are not limited to: the amount of biological sample required for a particular method, the sensitivity and coverage of detecting and quantifying differentially expressed transcripts, the number of samples to be processed, and the propensity for scale-up to industrial high-throughput. There is a wide range (from tens to hundreds of nanograms) in the amount of material (mRNA, cDNA, cRNA, etc.) required for each of the various methods. This may be an important concern if one wishes to quantify transcript abundance from a source that is limiting (e.g., tissue biopsy). It is essential to have a technology sensitive enough to detect mRNAs present as rare as one copy per 100,000 mRNAs. Achieving this level of sensitivity is necessary given the fact that 90 to 95% of all mRNA species are present at five or fewer copies per cell.\textsuperscript{1,2,3} Equally significant is the coverage or percentage of all possible transcripts assayed by a technology. The strengths and weaknesses of a gene expression technology can be further evaluated via the following elements: resolution, false-positive rates, and false-negative rates. Resolution is a measure of the ability of a method to distinguish any gene from its nearest neighbor — both in terms of sequence similarity (e.g., splice variants, gene families) and gene expression profile (e.g., co-regulated genes). The false
positive rate is the amount of genes incorrectly identified as differentially expressed. The false negative rate is the number of genes that are truly differentially expressed but not detected by the technology. All of the differential gene expression technologies are grouped into two major categories, closed and open architecture systems, and are summarized below. The aforementioned variables need to be considered when selecting the most effective and appropriate application of one of these two main classes of DGE technologies.

A. CLOSED SYSTEMS

Closed architecture systems are intrinsically constrained; one can only interrogate a finite set of genes — ones that are already known. Coverage is dependent upon the completeness of the knowledge of the given genome being assessed and the ability to represent it on a closed platform; hence restricting utility to querying of the most thoroughly sequenced and annotated species or biosystems. To be effective, closed systems must anticipate transcriptome (the context dependent mRNA profile) variations such as splice variants, polymorphisms, and RNA editing, etc., to maximize coverage. Closed systems play a fundamental role in the high-throughput analysis that follows the characterization of a transcriptome, for example, when a defined set of genes are to be interrogated under a large number of proscribed conditions.

Currently, the most widespread methods of closed systems analysis are DNA microarrays and real-time quantitative polymerase chain reaction (RTQ-PCR) (Table 25.1). DNA microarrays or “chips” are created in a reproducible pattern of hundreds or thousands of known sequences of cDNAs or oligonucleotides adhered to a solid support such as glass and are designed to simultaneously assay the expression of these genes in a single hybridization procedure. The hybridization of fluorescently labeled DNA or RNA complements is detected by laser scanning. Differences in labeling intensity indicate differences in quantitative gene expression. Real-time quantitative PCR, as its name implies, attempts to quantify the amount of a given transcript by measuring the accumulation of a PCR product over time using gene-specific primers and fluorescent probes (see Chapter 22 by Stephen A. Bustin in this book). RTQ-PCR provides greater sensitivity than microarrays, successfully amplifying transcripts from as little as 500 fg of starting RNA, and is therefore a popular technology for confirming differentially expressed genes identified using other platforms. RTQ-PCR can provide quantitative expression profiles of known transcripts across hundreds or

| TABLE 25.1 | Methods for High-Throughput Analysis of Differential Gene Expression (DGE) |
| Technology | Coverage (%) | Sensitivity | References | Company/Websites of Interest |
| DNA Microarrays | Variable | 1:300,000 | 6–8,4,5 | Affymetrix/www.affymetrix.com, www.microarrays.org |
| RTQ-PCR | 100 | 1:300,000 | 4,9 | Applied Biosystems/www.appliedbiosystems.com |
| GeneCalling | 95 | 1:125,000 | 4,5,15 | CuraGen Corporation/www.curagen.com |
| READS | >95 | <1:100,000 | 4,5,14 | GeneLogic/www.genelogic.com |
| SAGE | >92 | <1:10,000 | 2,4,5,10 | Genzyme/www.genzyme.com, www.sagenet.org |
| TOGA | 98 | <1:100,000 | 4,5,13 | Digi tal Gene Technologies/www.dgt.com |

* Patents covering open system technologies include GeneCalling (USPTO 5,972,693), READS (USPTO 5,712,126), SAGE (USPTO 5,695,937), and TOGA (USPTO 5,459,037).

* Coverage refers to the % of a cell’s/tissue’s expressed genes detected (modulated and unchanged), directly correlated to the number of iterations performed.

* Sensitivity refers to the number of copies of a single gene as a fraction of all expressed genes in a cell/tissue in order for that gene to be assayable by the technology.
thousands of samples simultaneously. Similarly, microarrays are relatively fast and extremely sensitive, and consume a minimum of reagent and sample due to miniaturization (see Chapter 13 in this book). However, they require continual updating and a catena of replicated experiments to incorporate new sequence information. Also of concern are the lot-to-lot variability in manufactured chips, the number of genes and gene variants that can be analyzed using a single chip and the finite number of splice variants, polymorphisms, insertions, deletions, etc., that can be resolved.

B. OPEN SYSTEMS

In addition to providing global views of gene expression similar to their closed system counterparts, open architecture systems are particularly suited for novel gene discovery. Open system methods are perceived to be more time consuming than closed systems. However, the ability to discover truly novel transcripts (genes, expressed single nucleotide polymorphisms, or cSNPs, alternative splicing, RNA editing, etc.), the serendipitous association of a gene or gene family with a disease or treatment, and the ability to monitor the coordinated behavior and expression of the entire transcriptome in a single cell or tissue simultaneously, provides significant advantage.

Current popular open technologies include serial analysis of gene expression (SAGE), total gene expression analysis (TOGA™), restriction enzyme analysis of differentially expressed sequences (READS™), and GeneCalling® complemented with SeqCalling™ when the species or tissue is not well represented in available sequence databases (Table 25.1). SAGE quantifies the expression levels of genes (including rare genes) by isolating short fragments (called tags) from the region close to the 3’ end of the cDNA following biotin-labeling of the poly-A tail of assayed messages and digestion with restriction enzyme. The tags are linked, sequenced, and a software package counts their relative frequency, forming a library, which can be used to assay the differences in expression between cells.

Although the use of SAGE as a tool for gene discovery and gene expression analysis has grown steadily over the last several years, including its use for the creation of a public database for gene expression in human cancers, a drawback of SAGE includes its dependence upon additional technology (such as Northern blots or RTQ-PCR) to confirm gene identity and expression differences. Similar to SAGE, TOGA also captures the 3’ end of the cDNA, which is then cut with restriction enzymes, cloned, and transcribed in vitro, followed by reverse transcription and two rounds of PCR. The PCR products are separated by electrophoresis and identified. Confirmation is via further PCR reactions. A sensitivity of detecting an mRNA present at 1 in 100,000 is reported and potential transcriptome coverage of 98% is possible with up to four iterations of the method. READS is also based upon restriction enzyme digests of double-stranded cDNA, generating only one tag for each gene fragment. The farthest 3’-end fragment including part of the poly-A tail is amplified and differentially expressed genes are then identified by differences in gel band intensities, followed by excision, cloning, and sequencing to confirm their identity. The coverage (>95%) and sensitivity (>1:100,000) are similar to that of the other open systems described.

Perhaps the most comprehensive coverage from an open architecture system comes from GeneCalling, an expression profiling technique that culls information delineated by restriction sites and electrophoretic mobility of cDNA-derived fragments. This technology, and its companion technology SeqCalling (described below), covers approximately 95% of an expressed genome. Expressed genes (mRNAs) are isolated from each sample and processed into cDNA fragments following a dual restriction digest. The cDNA fragments are end-labeled, PCR amplified, resolved by gel electrophoresis, and sized with an accuracy of 0.2 bp using capillary electrophoresis. Comparisons are made between fragments present in each sample (e.g., disease vs. normal) and expression differences are measured in silico using an integrated bioinformatics platform. Each differentially expressed fragment is queried against a species-specific database for gene identification; the known restriction enzyme pair sequences (six nucleotides at either end) along with the measured length of the cDNA fragment are adequate to predict or “call” proposed gene designations.
(or novelty). A probability score is associated with each GeneCall. Novel gene fragments can be eluted from agarose gels, sequenced, and cloned. The association of a modulated cDNA fragment to a known gene sequence is confirmed by a sequence-specific competitive PCR reaction known as oligonucleotide poisoning. In this process, the initial GeneCalling reactions are repeated in the presence of an unlabeled PCR primer consisting of one of the two 6-bp restriction enzyme sites of the predicted match plus additional nucleotides complementary to the corresponding region from the candidate gene. If the custom unlabeled primer competes with the standard labeled primer in the amplification reaction, then the selected cDNA fragment will not be detected in the repeat reaction. The fragment will be considered “poisoned,” and the association of the cDNA fragment to the candidate gene will be confirmed. This confirmation approach is contained within the GeneCalling platform, which is one of the features that distinguish it from the other open system technologies currently available. Generation of multiple fragments per gene increases the significance of initial GeneCall designations and reduces the time required for gene confirmation. GeneCalling is enhanced and expedited through complementation with SeqCalling when the species of interest is not well represented in sequence databases such as GenBank. SeqCalling is a validated high-throughput industrial-scale technology to generate comprehensive expression databases for any cell type, tissue, or organism thereby speeding discovery, providing maximum coverage of gene-coding regions, and enabling identification of rare transcripts present (down to 1 in $10^6$ copies) as well as single-nucleotide polymorphisms or cSNPs. Expressed genes are isolated and processed into cDNA fragments using the same 96 pairs of restriction enzymes that are used in the GeneCalling process. The fragments are size fractionated using high-resolution electrophoresis, cloned into standard vectors, PCR-amplified, and subjected to the same precise sizing using capillary electrophoresis as in the GeneCalling process. Clones within 0.2 bp are grouped by size and a representative from each size-group is sequenced. Following sequencing chemistry and electrophoresis, the sequences are fed into the bioinformatics process for sequence assembly and annotation to generate a custom database that is normalized and biased toward gene-coding regions providing efficient mapping of differentially expressed gene fragments to database sequences within a single platform.15

Industrial-scale open and closed systems can be used together to frame further, directed investigation of genes and their roles in biological pathways. Ideally, open systems can provide the context for specific application of a closed system. In biotechnology, these technologies are typically used to identify genes critical for the onset or progression of disease, identify potential drug targets, and evaluate the safety of potential therapeutics. The successful integration of multiple gene expression technologies, including both open and closed platforms, can streamline the gene discovery and drug development process and save time and research dollars by leveraging the inherent advantages of each technology. The remaining sections of this chapter are intended to provide an overview of how DGE technologies are applied to a specific area of research — drug target discovery.

II. APPLICATIONS TO DRUG TARGET DISCOVERY

A. FUNCTIONAL GENOMICS — GENE TARGET DISCOVERY

A generalized strategy for identifying disease-associated genes is outlined in Figure 25.1. It should be noted that this figure is greatly simplified, and does not reflect the integration of multiple technologies used in the pharmaceutical industry to identify and qualify drug targets independent of and complementary to expression profiling.

A fundamental assumption is that disease results from dysregulated gene expression combined with changes in normal cellular behavior. It then follows that discovering those genes that are differentially expressed in the context of disease, whether in an animal model, patient population, or an in vitro system that allows for a “disease vs. healthy” comparison, will reveal potential drug targets for the treatment of the disease. Application of an open profiling system will enable the identification of nearly all differentially expressed genes, spanning the continuum from genes with
a well-established link to the disease to those that are completely novel (Figure 25.1, left panel). The genes thus identified are the starting point for pharmaceutical and diagnostic research development. A priori, it may not be known whether the dysregulation of a specific gene(s) (depicted as Gene X in Figure 25.1) is causal or results from the disease state. Ultimately, downstream validation technologies, such as gene knockouts in transgenic models, cellular assays, protein biochemistry, and additional wet-lab validations will provide the necessary insight into the nature of the disease association that will be critical to the success of Gene X (or Protein X) progressing along the path toward the clinic.

“Genomic mining” provides a second method for identifying genes associated with disease and potential drug targets. Genomic mining attempts to extract biological meaning from novel sequences by functionally classifying them based on homology (sequence or protein structure) to known genes or gene families. Genes that fall into biological classes suitable for therapeutic intervention may
be pursued as potential drug targets. One possible approach is to use a closed architecture system such as RTQ-PCR to ascertain the expression profile of these “novel” genes across many tissues of interest (healthy and disease) and to triage or prioritize them as potential drug targets (Figure 25.1, right panel). Again, downstream technologies are essential in defining disease association and validating genes as pharmaceutical targets or drug candidates.

B. TOXICOGENOMICS — PREDICTING TOXICITY OF DRUG CANDIDATES

Toxicogenomics, the merging of genomics technologies with traditional toxicology, is a powerful new application for open expression profiling. The underlying assumption is that gene expression must be altered prior to and during the manifestation of toxicity and thereby can be predictive of toxicity. The strategy is to determine the expression profiles of both toxic and nontoxic compounds in model systems (such as rat or mouse) in order to identify a set of core marker genes that are dysregulated in target/sensitive tissues. These genes have predictive value, assuming that unknown compounds (e.g., potential drug candidates) that elicit a similar transcription profile produce similar toxicological outcomes. Most likely one of the tissues to be examined is the liver — the organ where the majority of drugs are metabolized and where toxicity is frequently observed in clinical trials. The use of open architecture profiling can validate previously known genes implicated in toxicity, as well as uncover both novel genes and known genes that may not have been previously associated with toxicity — especially if toxicity is being assessed in species that are not well represented in sequence databases (e.g., dogs, monkeys, rabbits, etc.). Furthermore, genes dysregulated in the toxic response can determine the likely mode of action of the toxic compound by highlighting key biochemical pathways. These pathways, and therefore the sets of genes that can be used to predict toxicity, will most likely be distinct (at least partially) for different toxicological states and different sets of compounds. Technologies that are amenable to high-throughput expression profiling (e.g., GeneCalling and microarrays) are being utilized to generate large databases of compound and toxicity-specific expression data. These same methods can be used to rapidly generate gene expression patterns for large numbers of drug candidates. Bioinformatics platforms that can integrate and organize data from a variety of sources and platforms, perform in silico (computer modeled) comparisons of expression data sets to identify differentially expressed genes, and reliably predict whether a compound will likely be toxic or not will be necessary to reap the full benefits offered by these technologies.

III. CONCLUSION

Genomics technologies have arisen out of necessity. The large influx of genome sequence data from the Human Genome Project (HGP) and from genome sequencing projects from a variety of organisms have resulted in greater demand by the end users of such data to have the resources to effectively organize, integrate, and analyze genome sequence information. However, mere sequence data is not enough to understand the complexity of the expressed genome — the transcriptome — both in terms of the specific genes present in the genome and also the ways in which the expression and regulation of these genes influence young vs. old, brain vs. muscle, normal vs. disease, benign vs. toxic, etc. Adding functional annotation to novel genes will be necessary to link sequence and expression data to biological meaning. The aforementioned DGE technologies have emerged as powerful tools to address these very questions. Those that are efficient, sensitive, and open to novel gene discovery hold the most promise.

ACKNOWLEDGMENTS

The authors wish to thank Drs. Dave Lewin, Cynthia Green, Pascal Bouffard, Joseph Rinninger, and Partha Seshaih for their critical review of this chapter.
REFERENCES

I. INTRODUCTION

Genome mapping projects have provided a wealth of information concerning the genomic sequence of higher organisms. This information can be used to investigate the underlying genetic changes giving rise to a variety of normal, developmental, and disease states. It is important, however, to study how genes work in concert rather than in isolation. Several methods are capable of determining global gene expression changes, including subtractive hybridization,\(^1\) differential display,\(^2\) and cDNA microarray hybridization.\(^3\) However, the numbers of different transcripts that can be analyzed with subtractive hybridization and differential display are limited. Microarray technology has the potential to examine expression of a larger number of genes but requires some a priori decision on what sequences should be examined. The SAGE method, originally described in 1995 by Velculescu et al.,\(^4\) has all the advantages of an “open system” (see overview Chapter 25 by Timothy Lohret and Karen Kelly); it allows quantitative and qualitative analysis of a large number of transcripts without prior knowledge concerning about which might be affected.\(^5\)
SAGE generates short sequence tags (9 to 10 base pairs) from the cDNA pool of the sample. The tags are generated as dimers or ditags and are ligated together end to end to form concatemers, which are then cloned. Large-scale sequencing of many SAGE clones allows over 30 individual tags to be read from each lane of an automated sequencing gel. The abundance of a particular tag detected after sequencing many clones relates directly to the expression level of the gene from which it is derived. Thus, serial analysis of many thousands of gene-specific tags gives rise to an expression profile of the sample tissue.

SAGE applications span a wide range but fall into two main types. The first uses SAGE for the global characterizations of particular tissues or organisms. Together with the completed sequence of yeast genome, for example, these data allowed to identify physical regions of transcriptional activity. The second type of SAGE applications are techniques for comparing cells or samples. For example, Zhang et al. performed detailed expression analyses of carefully matched tissues to study gene expression changes in gastrointestinal tumors. Although there was extensive similarity between the samples, more than 500 transcripts were expressed at significantly different levels. This showed the underlying differences between normal and neoplastic cells and led to the identification of potential diagnostic or prognostic markers. The effect of the tumor suppressor gene p53 on the expression of other genes was studied in rat and human cell lines. In the rat study, the SAGE analysis of approximately 15,000 genes revealed that the expression of 14 genes was dependent on functional p53 protein whereas the expression of three genes was significantly higher in cells producing nonfunctional p53 protein. This study described the use of SAGE in evaluating global transcriptional responses of cell populations.

In the human study, many genes found to be markedly up-regulated in response to p53 induction were predicted to code for proteins that could either generate oxidative stress or responded to it. Collectively the gene expression changes led to a novel pathway through which p53 results in apoptosis.

II. MATERIALS AND METHODS

A. REAGENTS

The following kits were commonly used:

1. Preparation of starting materials: RNA and cDNA.
   a. Total RNA: To TALLY RNA — Total RNA Isolation Kit, Ambion, catalog #1910.
   b. mRNA: MicroPoly (A) Pure — mRNA Isolation Kit, Ambion, catalog #1918.
   c. cDNA synthesis system: GibcoBRL, catalog #18267-013.

2. Cloning concatemers and sequencing.
   a. pZero cloning kit (Invitrogen, catalog #K2500-01).
   b. SphI restriction enzyme (NEB catalog #182S).
   c. Purification of PCR reactions: Microcon 100 columns (Amicon, catalog #28142), QIAquick 8 PCR purification kits (QIAGEN, catalog #28142/28144).

3. Analysis of SAGE data is carried out using specialist SAGE software, which is freely available to academic users. Details are available on the SAGE Web page.

B. PREPARATION OF SAGE LIBRARIES

The necessary experimental detail required to perform SAGE cannot be published here in order to comply with conditions of the licensor, Genzyme Molecular Oncology. The Detailed Protocol is, however, freely available to academic users and can be obtained through the SAGE Web page. This way, the terms of the license can be adhered to and the Web page will ensure that the protocol is constantly updated.

www.taq.ir
It is essential to have high quality reagents in order to obtain high quality SAGE libraries. The many SAGE reagents and linker molecules used for SAGE are now available in kit form (I-SAGE™) available from Invitrogen.

An overview of the SAGE protocol is shown in Figure 26.1. The method is described in more detail below and in Figure 26.2.

1. Preparation of Biotinylated 3’ cDNA

It is important in generating good SAGE libraries with large clone inserts to have the best quality starting material possible. Any standard protocol or commercial kit can be used to prepare the RNA and cDNA samples. Those detailed above have been found to be reliable.

High quality linkers are crucial to several steps in the SAGE method. Linker oligos and the biotinylated oligo-dT should be obtained gel-purified.

Figure 26.2A shows how the cDNA sample is bound to streptavidin coated magnetic beads and digested with the anchoring enzyme — NlaIII is most commonly used. The captured 3’ cDNA fragment population is then divided in half. Each is then ligated to SAGE linkers A or B. A variety of linkers can be used at this point in SAGE. Linkers must contain the appropriate anchoring enzyme overhang, a restriction site for a type IIS enzyme (tagging enzyme), and a priming site for PCR amplification.

2. Formation of the SAGE Tag

Figure 26.2B shows how the tagging enzyme BsmFI is used to cleave the novel stretch of 3’ cDNA away from the streptavidin bead. BsmFI is a type IIS restriction endonuclease[12,13] that cleaves a defined distance away from its recognition site. The SAGE linker contains the recognition site for BsmFI. Digesting with this enzyme releases the linker and cDNA from the streptavidin bead. The linker is now attached to a short stretch of novel 3’ cDNA. This forms the SAGE tag.
3. Formation of the SAGE Ditag

Figure 26.2C shows the ligation of the tag A population to the tag B population to form ditags. The primers contained within the linkers are then used to amplify the ditag sample. After PCR conditions have been optimized, large-scale PCR (100 to 200, 50µl reactions) can be performed for high yields of the 102-bp ditag/linker molecule.
4. Formation of Concatemers and Clones

Figure 26.2D describes restriction enzyme digestion of the 102 bp ditag molecule with the restriction enzyme \(Nla\)III. This releases the 26 bp ditags and 40 bp linkers. Both these molecules have the same sticky ends, which presents a potential problem for subsequent steps. The linker molecules have to be removed in order to avoid the ligation of a linker molecule to a growing ditag concatemer; the linker molecule does not have correct sticky ends for the cloning step and will effectively poison the concatemer.

In the detailed protocol this purification entailed a gel purification of the ditag molecule away from the smaller, unwanted linkers. An enhanced ditag purification method to address the problem described above has been published and compared with the alternative method of the Detailed Protocol. Details of this modification can be found elsewhere. The purified ditag molecules are then ligated to form end-to-end concatemers, which are cloned into a suitable vector, commonly pZero as described in Materials and Methods.

www.taq.ir
5. Analysis of SAGE Clone Inserts

The clone inserts are analyzed by PCR amplification and the insert sizes checked by agarose gel electrophoresis, as shown in Figure 26.3. The clones containing large insert sizes (commonly >600 bp) were selected and purified for sequencing. For the cloning system described here, a clone insert consists of ~226 bp of vector sequence plus concatenated ditags each of which is 26 bp. Each ditag represents two tags. Therefore, a clone of 616 bp equates to 30 SAGE tags — 226 bp of vector sequence plus 15 × 26 bp ditags.

5. Analysis of SAGE Clone Inserts

The clone inserts are analyzed by PCR amplification and the insert sizes checked by agarose gel electrophoresis, as shown in Figure 26.3. The clones containing large insert sizes (commonly >600 bp) were selected and purified for sequencing. Typical SAGE clone sequence is also shown in Figure 26.3. The NlaIII sites remaining between the ditags serve as punctuation for the concatemer and determine where a ditag starts and finishes.

Sequencing can be performed manually or by automated sequencing. Even in the case of small-scale SAGE projects, the sequencing effort is high (the generation of a minimum of 10,000 tags per library would require sequencing over 350 clones of the size described here). Automated sequencing of large-insert SAGE clones is therefore recommended to minimize this workload. ABI 377 automated DNA sequencers (Applied Biosystems, Foster City, CA) have been commonly used here.

C. Interpretation of the Data

The SAGE sequence data resulting from automated sequencing can be input directly into the SAGE analysis software. The SAGE software then creates project files for the analysis of clone sequence after defining enzymes used, tag, and ditag lengths. The software identifies the anchoring enzyme sites between ditags and records the tags and duplicate ditags found.

After the input of many clones, the project files of SAGE tags can be analyzed to determine the expression profile of the starting tissue. A SAGE tag project is also created from the sequences in Genbank in order to find a tag in the correct position in the transcript depending on which anchoring and tagging enzymes are used. The SAGE tags can then be compared with the Genbank library in order to detect matches for a particular tag and to determine the gene or EST from which it was derived.

A variety of statistical methods are currently being used in order to determine whether an expression difference is significant given other parameters of the SAGE tag project. The SAGE...
software incorporates facilities to perform significance calculations. The SAGE software files can also be exported into MS Excel or MS Access in order to further analyze the data.

Gene identities were defined from tag sequence using the SAGE tag to gene facility maintained by the National Institutes of Health\textsuperscript{16} and Tagsorter\textsuperscript{TM,17} Positive assignments were made on the presence of a poly-A tail and the tag being detected within the transcript at the correct position, namely adjacent to the 3'-most \textit{NlaIII} site.

III. DISCUSSION

A. Applications of SAGE

Recent examples of the use of SAGE in the analysis of transcript profiles of an organism or tissue include a study of the human epidermis in which over 15,000 SAGE tags were derived from purified keratinocytes.\textsuperscript{18} A much larger study, comprising over 40,000 SAGE derived from the mouse neocortex\textsuperscript{19} enabled the selection of candidate genes to be used in subsequent \textit{in situ} hybridization experiments to reveal spatial expression patterns in this tissue.

In contrast, large-scale comparative studies have been carried out in plants in order to study differential gene expression of lignifying xylem.\textsuperscript{20} This study resulted in the analysis of over 150,000 SAGE tags, representing over 25,000 expressed genes, currently the most extensive analysis of its kind in a higher plant. Human cardiac cells were used for a comparative SAGE study in order to determine the complex cellular responses involved in hypoxia.\textsuperscript{21} The breast cancer promoting effects of estrogen on gene expression have also been compared with the chemopreventative effects of tamoxifen. This was carried out in a comparative SAGE study using differential treatment of the ZR75-1 estrogen-dependent breast cancer cell line.\textsuperscript{22}

B. Development of the SAGE Technology

The choice of starting material is an important consideration when embarking on a SAGE analysis. In order to directly compare two transcript profiles, the source mRNA has to be matched suitably to take into account variation, for example, between individual tissue samples. To this end cell lines provide plenty of material for SAGE. Comparisons of lines with and without the inclusion of a transfected gene provide a constant background upon which to view gene expression changes occurring as a consequence of the introduced gene. Many such SAGE libraries have been constructed and the tag libraries of the tissues concerned are posted on the SAGE website. The accumulation of such data provides valuable additional information to gene identification studies.\textsuperscript{23–25} As SAGE technology improves it is possible to target defined cell types in order to greatly refine the starting material and hence accumulate the resulting libraries of a SAGE analysis. The SAGE method has seen developments in the technology resulting in SAGE lite,\textsuperscript{26} Micro SAGE,\textsuperscript{27} and SADE,\textsuperscript{28} allowing smaller amounts of starting material to be used to make libraries. These developments widen the possible applications of the technique to comparisons between cell types of limited availability\textsuperscript{29} and allow the use of precision methods such as laser microdissection to procure specific cell samples.\textsuperscript{28}

C. Using SAGE to Annotate the Human Genome

The computational methods used to identify putative gene sequences from the large-scale sequencing of the human genome have limitations as they are unable to provide definitive evidence on whether a hypothetical gene is actually expressed. In their recently published paper, Saha et al.\textsuperscript{30} addressed this problem and used SAGE data to corroborate the identification of genes from the efforts of the human genome project. This paper describes the use of LongSAGE. In this modification of the SAGE method, an alternative type IIS restriction endonuclease \textit{MmeI} is used to generate the SAGE tags. The modifications result in the production of longer transcript tags typically

www.taqr.ir
comprising 17 bp of novel 3’ cDNA sequence. The matching of such longer tags to the genomic sequence allows precise localization of the tags within the genome. The comparison of a tag location with the position of a previously annotated gene provides additional expression evidence for previously predicted genes. This can also lead to the identification of novel internal exons and previously uncharacterized genes. This is a unique demonstration of the power of the SAGE technique in the evaluation of novel gene sequences.

ACKNOWLEDGMENTS

Jill Powell was funded by Cancer Research U.K. at the Richard Dimbleby department of Cancer Research, St Thomas Hospital, Lambeth Palace Road, London, SE1 7EH.

REFERENCES

9. A detailed SAGE protocol and clone analysis software can be obtained via the SAGE Web page (http://www.sagenet.org/).
10. Genzyme Molecular Oncology (One Mountain Road, P.O. Box 9322, Framingham, MA).

www.taq.ir
The Detection of Differential Gene Expression Changes Using SAGE


I. Introduction ................................................................................................................. 277
   A. Definition and Theoretical Aspects ......................................................................... 277
   B. Applications in the Area of Infectious Diseases ....................................................... 278
   C. Optimization of DD .................................................................................................... 279
II. Materials and Methods .................................................................................................. 280
   A. Preparation of Total Cellular RNA ........................................................................... 280
   B. Reverse Transcription ............................................................................................... 280
   C. Differential Display and Isolation of DNA Bands for Sequence Analysis .............. 280
   D. Identification of Genes ............................................................................................. 281
   E. Semiquantitative RT-PCR and Real-Time Sybr Green® RT-PCR .................................. 281
III. Differentially Displayed Genes During Dengue Virus Infection .............................. 282
IV. Conclusions ................................................................................................................. 283
Acknowledgments ................................................................................................................ 283
References .......................................................................................................................... 284

I. INTRODUCTION

A. Definition and Theoretical Aspects

Differential display (DD) is a powerful method used to detect gene expression in a number of biological systems to distinguish differential expression of mRNAs. DD results are easily reproduced, do not require enzymatic restriction, enzyme digestions, and ligations or cloning, and require only very small amounts of starting total RNA (from 20 ng to 200 ng) which results in a less labor-intensive method when compared with similar approaches.\textsuperscript{1–3} Gene fragments from differentially expressed genes can be excised from the DD gel, identified, and used to prepare gene tags for the study of gene expression levels, using semiquantitative and quantitative PCR.

Figure 27.1 shows the basic DD protocol. Various parameters in the PCR can be altered to increase the number of displayed bands on a gel (Table 27.1). The number of bands obtained per cDNA pool will vary with each random primer used. We have defined an equation (Equation 1, see box) to help understand the relationship between length of the primers and temperature of annealing used in the reaction and the probability to hybridize to a given RNA pool.

Experimentally, the optimal temperature for DD using a 13-mer random primer is between 50°C and 40°C.\textsuperscript{4} The signals obtained in DD need further confirmation by more quantitative and specific techniques. Quantitative RT-PCR, Northern blots, reverse Northern blots, macroarrays, or
microchips are reasonable possibilities. We have confirmed that the data obtained by DD in the present study is reproducible using RT-PCR, Northern blots, and microchip Affymetrix analysis using H133A human chips. Here we will be presenting such data and will give experimental details focusing on the DD part of it.

DD is defined by the fact that no \textit{a priori} comprehensive knowledge of the transcriptome is necessary, hence the field of discovery is open and therefore belongs to the category of “open” architecture system (see the overview Chapter 25 by Timothy A. Lohret and Karen M. Kelly). DD still is the most widely used method for expression analysis because it can be performed in any laboratory equipped with standard molecular biology reagents and instrumentation, and, in its most basic form, the need for advanced bioinformatics is minimal.

**B. Applications in the Area of Infectious Diseases**

DD has been applied to studying the cellular response to a large number of pathogenic organisms as well as to the study of the pathogens themselves. Modifications for the study of non-eukaryotic cells, mainly by arbitrarily primed PCR where both primers anneal randomly in the genome, have been proposed for prokaryotic cells. Examples are \textit{Mycobacteria tuberculosis},\textsuperscript{5} other bacteria,\textsuperscript{6,7}
and *Legionella pneumophila*. Nonmammalian cells have also been successfully studied — for example, plants, *Schistosoma* sp. and *Biomphalaria glabrata* vector, candida, the fungus *Leptosphaeria maculans*, *Plasmodium falciparum*, and vector arthropods.

DD has also been used successfully for studying the molecular mechanisms of antipathogen responses by host cells. Examples are infection with Cytomegalovirus, enterovirus, and coxsackievirus.

In the last part of this chapter we will demonstrate how DD was used to study the response of human vascular endothelial cells (HUVECs) to dengue virus 2 New Guinea C (NGC) infection in vitro. In this study, DD analysis provided specific information on at least three cellular pathways involved in the response to the infection. As a general strategy, we propose to provide a quantitative description of the behavior of genes during pathogenesis on a reduced scale and complement the discoveries using DD with hypothesis-driven experiments.

### C. Optimization of DD

We found that Sensiscript® RT (Qiagen) produced DD with a higher sensitivity as compared to other reverse transcriptases like MMLV and Superscript II (Life Sciences). The sensitivity of DD can also be improved by using $^{33}$P, which has a longer half-life than $^{32}$P and lower energy $\beta$ emission. Consequently, it has a favorable background-to-noise ratio and does not require special readers. False positives are a common problem with DD but can be avoided by adopting the following precautions:

1. Utilize homogenous cell preparations. Avoid the utilization of complex mix of cells or tissues from different origin.
2. Utilize an RNA extraction method that is suitable for the amount of starting material. If the cells exceed $1 \times 10^7$, the most efficient way to obtain total RNA is extraction with an organic phase such as Trizol (Life Sciences). For samples containing less than $10^7$ cells, it is best to avoid the use of organic solvent extraction and alcohol precipitation. Avoid the use of DNaseI clean kits that require reprecipitation of the purified RNA. Instead, utilize an affinity column for RNA extraction that includes a final step with DNaseI treatment. Always use DEPC treated water to elute the RNA. Assess RNA

---

**TABLE 27.1**

**Comparison of Traditional and Modified DD**

<table>
<thead>
<tr>
<th>Traditional Method</th>
<th>Differential Display</th>
<th>Modified Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reverse transcription using MMLV or AMV reverse transcriptase. Start with 0.2-1 μg total RNA.</td>
<td>Reverse transcription using Sensiscript. Starts with 20–100 ng total RNA.</td>
<td>Reverse transcription using Sensiscript. Starts with 20–100 ng total RNA.</td>
</tr>
<tr>
<td>cDNA bands of interest are excised in water and purified from acrylamide.</td>
<td>cDNA bands of interest are excised and purified from acrylamide in the presence of salts (2xPCR buffer).</td>
<td>cDNA bands of interest are excised and purified from acrylamide in the presence of salts (2xPCR buffer).</td>
</tr>
<tr>
<td>PCR isolated cDNA fragment using same primer combination used for DD.</td>
<td>PCR isolated cDNA fragment using same primer combination used for DD.</td>
<td>PCR isolated cDNA fragment using same primer combination used for DD.</td>
</tr>
<tr>
<td>Confirm DD results by Northern and sequencing of clones.</td>
<td>Confirm DD results by Northern and sequencing of clones.</td>
<td>Confirm DD results by Northern and sequencing of clones.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sequences are queried against NCBI databases using Basic Local Alignment Search Tool (BLAST).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Confirm DD results using specific primers or probes by Real Time RT-PCR.</td>
</tr>
</tbody>
</table>

and *Legionella pneumophila*. Nonmammalian cells have also been successfully studied — for example, plants, *Schistosoma* sp. and *Biomphalaria glabrata* vector, candida, the fungus *Leptosphaeria maculans*, *Plasmodium falciparum*, and vector arthropods.
amounts by absorbance at 260 nm. Store RNA at −70°C at high concentration, if possible greater than 1 mg/ml.

3. DD should be performed preferentially with freshly prepared cDNA. Use high performance Sensiscript® or Omniscript® (Qiagen) reverse transcriptases. Do not exceed the amount of RNA recommended, usually between 20 ng and 200 ng, respectively, per reaction.

4. Run duplicates of cDNAs when tissues are the primary source of RNA. If a pathogen is present in the cellular preparation, include a cDNA that contains the genetic material of the pathogen, plus the untreated cellular RNA. This is very important when virus-infected cells are used and the virus has an RNA genome. Start the DD with an anchor primer containing C or G before the oligo dT.

5. Always run duplicates for the low-stringency PCR reactions per each cDNA utilized. Use hotstart Taq polymerase to avoid false priming and dATP$_{33}$P as the source of radioactive nucleotide.

6. When drying the gels, assure complete removal of the urea.

7. Mark the glass plates asymmetrically with $\beta$ fluorescent ink marks prior to exposure with the x-ray film. This will facilitate proper centering of the gel when cutting the bands.

8. Overlay the cut bands with 2 x PCR buffer instead of water.

9. Purify the eluted fragments with a PCR clean-up kit prior to sequencing of the bands. If core facilities are used for automated sequencing, sequencing reactions should be performed at a low stringency temperature (40°C). The DD method with modifications is summarized in Table 27.1.

II. MATERIALS AND METHODS

A. PREPARATION OF TOTAL CELLULAR RNA

Total cellular RNA was extracted from $2 \times 10^7$ uninfected or infected HUVECs using an RNA Extraction Kit plus DNaseI treatment (Qiagen, Germany). Typically, $1 \times 10^6$ cells produced 10 μg of total RNA. For smaller amounts of cells used, DNase I treated RNA was obtained from the Absolutely RNA™ kit (Cat. #400805, Stratagene). The RNA was resuspended in diethyl pyrocarbonate (DEPC)-treated dH$_2$O and was quantified by spectrophotometry (Bio Photometer, Eppendorf, Germany) at 260 nm. Infection and culture of HUVECs has been described by our laboratory in previous work.

B. REVERSE TRANSCRIPTION

RT was performed using Sensiscript® Reverse Transcriptase (Cat. #205211, Qiagen, Germany). cDNA was synthesized in 20 μl reactions containing 50 ng of HUVEC mRNA, RT buffer (25 mM Tris-Cl, pH 8.3, 37.2 mM KCl, 1.5 mM MgCl$_2$, 5 mM DTT) and 0.5 mM dNTP, 10 μmol/μl of H-T11C anchor primer (5'-AAGCTTTTTTTTTTTC-3') (GenHunter kit RNAimage), and 10 U/μl RNase inhibitor. Reactions were incubated at 37°C for 60 min and then at 95°C for 5 min.

C. DIFFERENTIAL DISPLAY AND ISOLATION OF DNA BANDS FOR SEQUENCE ANALYSIS

Differential Display was performed as described previously with some minor modifications. PCR amplification of cDNA was done according to the RNAimage protocol (Genhunter, Nashville, TN) using 10% of the RT reaction and 0.4 μl of Titanium™Taq DNA Polymerase (Clontech, Palo Alto, CA), 2 μl of Titanium Taq PCR buffer (final concentrations of 10 μM Tris-Cl, pH 8.4, 50 mM KCl, 1.5 mM MgCl$_2$, and 5 mM DTT), 200 nM primer mix, 200 nM d NTPs, and 0.2 μl of 2000 Ci/μmol of [α-33P]dCTP per 50-μl reaction in PCR-grade water. Conditions were 95°C for 1 min, then 40 cycles at 93°C for 30 sec, 40°C for 2 min, 68°C for 30 sec, and a final step at 68°C for 10 min. PCR reactions were electrophoresed on extended-format denaturing 6% polyacrylamide
gels using the programmable Genomyx LR DNA sequencer (Beckman Coulter, Columbia, MD). Alternatively, labeling with fluorochromes allows the digitalization of images with large-format scanners. The RNAspectra Kits (GeneHunter) feature a new generation of Differential Display technology based on fluorescent detection of PCR products instead of radioactivity.

The RNAimage (Genhunter, Nashville, TN) protocol with some modifications was followed to reamplify candidate bands observed on the DD gels. The bands were extracted from the denaturing polyacrylamide gel by soaking in 100 μl of 2x PCR buffer for 10 min. This supernatant was treated with 10 μl sodium acetate (3 M), 5 μl of glygen (10 mg/ml), and 450 μl of 100% ethanol, stored at –70°C for 30 min, and pelleted by centrifugation. The pellet was rinsed with 200 μl ethanol (80%), air dried, and dissolved in 10 μl dH2O. This extracted DNA was amplified (Genhunter).

The QIAquick Gel extraction kit protocol (Qiagen, Germany) was used to extract DNA from samples run on 1.5% agarose gels. Purified DNA was eluted in dH2O instead of Buffer EB (10 mM Tris-Cl, pH 8.5) and quantified by spectrophotometry using a BioPhotometer (Eppendorf, Germany) at 260 nm. DNA samples were sequenced in the University of Massachusetts medical school sequencing core facility, using the anchor primer or forward primers.

D. IDENTIFICATION OF GENES

The standard BLAST search using the website www.ncbi.nlm.gov was used to obtain information for each sequence. The advantages of using BLAST are the fast access to PubMed and MEDLINE, metadata information on the gene family and known cellular functions, taxonomic analysis, and links to UniGene System. Each UniGene cluster contains sequences that represent a unique gene as well as related information such as the tissue-specific expression and chromosome map location.

E. SEMIQUANTITATIVE RT-PCR AND REAL-TIME SYBR GREEN® RT-PCR

To confirm the DD results, we reverse transcribed 200 ng of total cellular RNA from infected and uninfected cells using the Omniscript Reverse Transcriptase RT-PCR kit (Cat. #205110, Qiagen, Germany) and 1 μM of oligo dT primer. For semiquantitative PCR, various quantities (0.5, 1 and 2 μl) of this freshly prepared cDNA were used together with 2.5 U of Taq DNA Polymerase (Clontech, Palo Alto, CA), and final concentrations of 100 μM of each dNTP and 0.1 μM of each primer in a total volume of 50 μl. The PCR program consisted of 94°C for 1 min, followed by a varying number of amplification cycles (58°C for 1 min, 72°C for 45 sec, and 94°C for 30 sec) and a final extension cycle of 58°C for 1 min and 72°C for 10 min. Results are shown in Table 27.2.

For real-time RT-PCR, cDNA was obtained as described above. An aliquot of 1 to 4 μl of cDNA was amplified by adding 15.75 μl PCR master mix [5 μl of 10X SYBR PCR Buffer, 6 μl of 25 mM MgCl2, 4 μl dNTPs (10 mM), 0.25 μl Amplitaq Gold (1.25 U) and 0.5 μl AmpErase

<table>
<thead>
<tr>
<th>Genes</th>
<th>Fold Induction SYBR®Green</th>
<th>Fold Induction RT-PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>h-IAP1</td>
<td>8.1</td>
<td>7.79</td>
</tr>
<tr>
<td>2’-5’OAS</td>
<td>72.85</td>
<td>2.65</td>
</tr>
<tr>
<td>2’-5’OAS like</td>
<td>67.7</td>
<td>3.55</td>
</tr>
<tr>
<td>ESDN</td>
<td>0.85</td>
<td>0.45</td>
</tr>
<tr>
<td>MxA</td>
<td>+++</td>
<td>11</td>
</tr>
<tr>
<td>RGS2</td>
<td>2.8/26.37</td>
<td>3.9</td>
</tr>
<tr>
<td>Gal-9</td>
<td>4.59</td>
<td>4.7</td>
</tr>
<tr>
<td>TNF α</td>
<td>3.76</td>
<td>3.82</td>
</tr>
<tr>
<td>IL1-β</td>
<td>40.77</td>
<td>7.79</td>
</tr>
</tbody>
</table>
UNG (0.5 U)], and deionized water ad 50 μl. Calibration curves containing five points ranging from 100 fg to 1 ng were used as standard for the experiment and L35A was used as an internal control for normalizing gene expression levels. Reactions were cycled at 50°C for 2 min and 95°C for 10 min for 40 cycles, after which at 95°C for 15 sec and at 60°C for 1 min were programmed in a PCR Signal Detection System 5700 (Applied Biosystems).

III. DIFFERENTIALLY DISPLAYED GENES DURING DENGUE VIRUS INFECTION

In the GenBank, eight known genes were found to have matches out of the 14 submitted sequences. The list of the genes is shown in Table 27.2. No false positives were detected. A total of 14 primer pair combinations (H-AP 1, 3, 4, 6, 8, 12, 15, 16, 19, 21, 22, 23, 41, and 42) were used. This represents around 57% efficiency of practical detection of identifiable GenBank sequences. Genes not identified in the GenBank are still of interest but we recommend to concentrate the work on identified sequences of DD fragments matching known genes.

The eight DD genes were confirmed by amplification protocols of RT-PCR. Of the eight differentially expressed genes, six are shown in Figure 27.2. We were able to identify com-

![Figure 27.2](image_url)

**FIGURE 27.2** Semiquantitative RT-PCR analysis. One of three semiquantitative RT-PCR analysis is shown for six out of eight genes detected by DD; uninfected (left) or virus infected cultures (right) at 48 h post infection are included in each panel. L35a was used as a control gene. In the PCR steps for all reactions, 0.5 μl, 1 μl, and 2 μl of cDNA of either uninfected or infected samples were used, except for L35a in which 0.25, 0.5, and 1 μl of cDNA were used in parallel for loading control.
components of at least three known signaling pathways, e.g., those used by IFN-α/β, IL-1β, and TNF-R. Figure 27.3 represent a model of the overall response of HUVECs to infection with dengue virus.

IV. CONCLUSIONS

The utilization of DD is a suitable way to explore gene expression. In HUVECs infected with dengue virus the cellular response found in vitro could potentially model in vivo infection of endothelium. We have shown that primary isolates of dengue virus obtained directly from patient samples can infect HUVECs, and therefore this infection model may reflect pathogenesis of the hemorrhagic disease. DD can provide information on genes associated with a complex biological system.

ACKNOWLEDGMENTS

The authors would like to acknowledge Miguel B. Bosch for the formulation of Equation 1, Kris Xhaja and Angela Ariza for excellent technical help in RT-PCR, and Jonathan H. Dinsmore and Alan L. Rothman for critical reading of the manuscript. This work was supported by grant R01 AI30624 from the National Institutes of Health (NIH). The opinions expressed herein are those of the authors and should not be construed as representing the official policies of the NIH.
Equation 1

\[
Q = \text{number of genes in a mammalian cell} \\
L = \text{average mRNA length} \\
p_{\text{p}} = \text{primer length} \\
nt = \text{number of base matches necessary for primer extension at a given annealing temperature} \\
p! (np - nt)! = \text{number of ways in which nt matches can occur}
\]

The probability of priming is

\[
p = \frac{1}{4}^{nt} \cdot \frac{3}{4}^{np - nt} \cdot \frac{L - np}{np} \cdot \frac{np!}{nt! (np - nt)!}
\]

and the estimated number of genes found by one DD reaction is \( Q \times Pr \)

If the number of expressed genes (\( Q \)) in a mammalian cell is near 20,000, the cDNA pools will contain 7,000 (the anchor primer ending in AT, GT, or CT is one out of three possible combinations). The average length (\( L \)) of a mRNA is likely to be about 2,705 bases (as suggested by The National Center of Biotechnology information http://www.ncbi.nlm.nih.gov/genome/guide/human/, excluding the poly-A tail length). With a 13 base-oligonucleotide (\( nt = 13 \)), and assuming that at 40°C primer extension requires only 7 out of 13 (\( np = 7 \)) nucleotides to match the target sequence, the probability of finding a cDNA species in the pool is 0.00186. Thus, the number of genes that are transcribed will be \( Q \times p = 130 \) genes which is very close to the number of bands we obtain per PCR reaction for primer combination. As we increase temperature of annealing in the DD PCR reaction, the number of bands obtained decreases. For example, if 8 instead of 7 nucleotides need to bind, the number of bands per reaction will be only 33.

REFERENCES


28 Restriction Endonucleolytic Analysis of Differentially Expressed Sequences: READS

Yerramilli V. B. K. Subrahmanyam
and Sherman M. Weissman

CONTENTS

I. Introduction................................................................................................................. 287
II. Materials and Methods..................................................................................................... 288
   A. Reagents ..................................................................................................................... 288
   B. Enzymes ...................................................................................................................... 289
   C. Oligonucleotides.......................................................................................................... 289
      1. T7-Sal-Oligo d(T)₆-V............................................................................................. 289
      2. READS PCR Primers ........................................................................................... 289
      3. Fly Adapters .......................................................................................................... 290
III. Experimental Protocol.................................................................................................... 290
   A. General Precautions.................................................................................................... 290
   B. Reagents To Be Prepared in Advance ....................................................................... 290
      1. Annealed Fly Adapter .......................................................................................... 290
      2. Labeled PCR Primer ............................................................................................. 290
      3. Labeled Molecular Weight Marker DNA ............................................................ 291
   C. cDNA Synthesis ........................................................................................................ 291
   D. Restriction Enzyme Digestion of the cDNA ............................................................. 292
   E. Ligation of Digested cDNA with Fly Adapter ......................................................... 292
   F. PCR Amplification .................................................................................................... 292
   G. Sequencing Gel Analysis .......................................................................................... 293
   H. Autoradiography ....................................................................................................... 293
   I. Recovery of cDNA Bands from the Gel and PCR Amplification ............................... 293
   J. Processing of the PCR-Amplified Samples for Sequence Analysis ........................... 294
   K. Data Documentation ................................................................................................ 294
   L. Technical Comments ............................................................................................... 295
IV. Discussion................................................................................................................... 296

References................................................................................................................................ 297

I. INTRODUCTION

Profiling of cDNA expression patterns has become a major tool of molecular biology and proved valuable in a range of problems from fundamental studies of biochemical mechanisms to the classification of human tumors.²⁻⁵ The largest number of studies has been performed by
hybridization of labeled cDNA to arrays of oligonucleotides or cDNA fragments and, with rare exceptions, the arrays represent only previously detected transcripts. Large-scale independent confirmation of these results has rarely been performed. In addition, very recent studies indicate that there may be much more transcriptional activity in cells than would be suggested from the known or predicted protein coding genes. Methods for gel-based display that are independent of prior knowledge about the transcripts have been described for expression analysis. These methods are applicable to any eukaryotic organism. Among the gel-based display approaches, READS™ (Restriction Endonucleolytic Analysis of Differentially Expressed Sequences) gel display process (as outlined in Figure 28.1) involves selective PCR amplification of the 3' end restriction fragments of the cDNAs based on the N1N2 nucleotides 5' to the poly-(A) tail of the message. As a result, this approach produces well-defined products from each polyadenylated transcript and, in general, produces only one major product per poly-(A) site on the transcript. It has a wide dynamic range and predicts changes in the relative levels of different RNA species. The approach does not require any specialized equipment and is suitable for use even in small laboratories. READS can be carried out even with limited amounts of total RNA and hence can be applied to small tissue biopsies. The present chapter provides a detailed protocol evolved over several years of experience that has given very consistent results in the authors’ laboratories.

II. MATERIALS AND METHODS

A. REAGENTS

Glycogen (20 mg/ml) from Boehringer-Mannheim, gel extraction and PCR purification kits from Qiagen, γ-[^32]P-Adenosine Triphosphate (ATP), 3000 Ci/mmol from Amersham Corporation, Quick
Restriction Endonucleolytic Analysis of Differentially Expressed Sequences: READS

Spin™ (Sephadex-G25) columns from Boehringer-Mannheim, Sequencing gel mix-6 and 1kbPlus ladder from Invitrogen, Seakem GTG agarose from FMC Corporation, Sigmacote from Sigma chemical company, Ammonium persulfate and \(N,N', N', N'\),-Tetramethylethelenediamine (TEMED) from Bio-Rad Laboratories and Bio Max, MR x-ray film from Kodak, and Radtape from Diversified Biotech were used.

Buffers for READS:
- 1 X Tris-borate-EDTA (1 X TBE) buffer: 90 mM Tris-borate (pH 8.3), 2 mM EDTA
- 1 X TE buffer: 10 mM Tris-HCl (pH 8.0), 1 mM EDTA
- 10 X annealing buffer: 100 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 1 M Sodium Chloride
- 5 X T4 Polynucleotide kinase exchange reaction buffer (Invitrogen): 250 mM Imidazole-HCl buffer (pH 6.4), 60 mM MgCl\(_2\), 5 mM 2-Mercaptoethanol, 350 \(\mu\)M Adenosine diphosphate
- 10 X T4 Polynucleotide kinase buffer (New England Biolabs): 700 mM Tris-HCl (pH 7.6), 100 mM MgCl\(_2\), 50 mM dithiothreitol
- Stop solution (USB Corporation): 95% v/v Formamide, 20 mM EDTA pH 8.0, 0.05% w/v Bromophenol blue, 0.05% w/v Xylene cyanol FF

B. ENZYMES

Superscript II Reverse Transcriptase (200 U/\(\mu\)l) along with 5 X first strand buffer [250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM MgCl\(_2\)] and 0.1 M Dithiothreitol (DTT), 5 X second strand buffer [100 mM Tris-HCl (pH 6.9), 23 mM MgCl\(_2\), 450 mM KCl, 0.75 mM \(\beta\)-NAD+ and 50 mM (NH\(_4\))\(_2\)SO\(_4\)], DNA Polymerase I (10 U/\(\mu\)l), *E. coli* DNA Ligase (10 U/\(\mu\)l) and Ribonuclease H (3 U/\(\mu\)l) from Invitrogen Corporation, RNase inhibitor (rRNasin, 40 U/\(\mu\)l) from Promega Corporation, Deoxyribonucleoside triphosphates (dNTPs, 100 mM), restriction enzymes, T4 DNA Ligase (400 U/\(\mu\)l) and T4 Polynucleotide Kinase (T4PNK, 10 U/\(\mu\)l) from New England Biolabs, AmpliTaq Gold (5 U/\(\mu\)l) with GeneAmp 10 X PCR buffer [100 mM Tris-HCl (pH 8.3), 500 mM KCl, 15 mM MgCl\(_2\), and 0.01% w/v Gelatin] from Applied Biosystems, shrimp alkaline phosphatase (1 U/\(\mu\)l) from Boehringer Mannheim, and Exonuclease I (10 U/\(\mu\)l) from USB Corporation were used.

C. OLIGONUCLEOTIDES

1. **T7-Sal-Oligo d(T)\(_{18}\)-V**

\[5'\text{-ACG TAA TAC GAC TCA CTA TAG GGC GAA TTG GGT CGA C (T)}_{18}\text{ V-3'}\]

2. **READS PCR Primers**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1-AA</td>
<td>5'-TGA AGC CGA GAC GTC GGT CG (T)(_{18}) AA-3'</td>
</tr>
<tr>
<td>P2-AC</td>
<td>5'-TGA AGC CGA GAC GTC GGT CG (T)(_{18}) AC-3'</td>
</tr>
<tr>
<td>P3-AG</td>
<td>5'-TGA AGC CGA GAC GTC GGT CG (T)(_{18}) AG-3'</td>
</tr>
<tr>
<td>P4-AT</td>
<td>5'-TGA AGC CGA GAC GTC GGT CG (T)(_{18}) AT-3'</td>
</tr>
<tr>
<td>Heel: P1-4</td>
<td>5'-TGA AGC CGA GAC GTC GGT CG-3'</td>
</tr>
<tr>
<td>P5-CA</td>
<td>5'-CTC TCA AGG ATC TTA CCG CT (T)(_{18}) CA-3'</td>
</tr>
<tr>
<td>P6-CC</td>
<td>5'-CTC TCA AGG ATC TTA CCG CT (T)(_{18}) CC-3'</td>
</tr>
<tr>
<td>P7-CG</td>
<td>5'-CTC TCA AGG ATC TTA CCG CT (T)(_{18}) CG-3'</td>
</tr>
<tr>
<td>P8-CT</td>
<td>5'-CTC TCA AGG ATC TTA CCG CT (T)(_{18}) CT-3'</td>
</tr>
<tr>
<td>Heel: P5-8</td>
<td>5'-CTC TCA AGG ATC TTA CCG CT-3'</td>
</tr>
</tbody>
</table>
3. Fly Adapters

These Y-shaped adapters have a region where adapter oligonucleotides FA-1 and FA-2 are annealed together and another region where these oligonucleotides are not complementary and do not anneal to each other, thus generating the Y-shaped fly adapter. The annealed fly adapter has a 5'-four-nucleotide overhang that is complementary to that generated by the restriction enzyme used to cut the cDNA. By changing the four nucleotides on the 5' end of FA-2 we can generate overhangs compatible with any restriction enzyme of interest. For the enzymes that produce a 3'-overhang (e.g., Pst-I), the FA-1 carries the corresponding overhang instead of FA-2.

FA-1 (Fly Adapter-1): 5'-TAG CGT CCG GCG CAG CGA CGG CCA G -3'
FA-2 (Fly Adapter-2): 5'-GAT CCT GGC CGT CGC TGT CTG TCG GCG C -3'

III. EXPERIMENTAL PROTOCOL

A. General Precautions

Use nuclease-free water and reagents. Plasticware should be of low adhesion quality to minimize losses. Clean all the pipettors with RNAse Zap wipes (Ambion) and use all precautions to avoid PCR contaminations. Always prepare sufficient mastermix for the number of tubes to be used and aliquot it accordingly to maintain consistency and reproducibility. While preparing the mastermix add the enzymes last and mix gently. Do not vortex. Isotopes and the isotope-containing gels are to be handled with the precautions recommended for radioactive materials.

B. Reagents To Be Prepared in Advance

1. Annealed Fly Adapter

Prepare the annealed adapters for all the enzymes planned for use in READS analysis, with the corresponding fly-adapter oligonucleotides FA-1 and FA-2 as described:

1. In a 1.5 ml tube, take 80 μl of nuclease free water, 10 μl of 10 X annealing buffer, and 5 μl (1 μg/μl) each of FA-1 and FA-2 adapter oligonucleotides. The final volume for the mixture is 100 μl, with an adapter concentration of 100 ng/μl.
2. Incubate the tube in a boiling water bath for 5 min, turn off the burner, and allow the annealing to take place till the bath reaches room temperature. Centrifuge briefly and store the fly adapter at -20°C.

2. Labeled PCR Primer

1. The FA-1 primer is 32P labeled and used for the PCR reactions. 2 μl of 1 μg/μl primer is labeled with γ-32P ATP in 20 μl of volume as described below:
   10 X T4 Polynucleotide kinase buffer 2 μl
   γ-32P ATP (3000 Ci/mmol) 15 μl (150 μCi)
2. Prewarm the tubes to 37°C, add 1 μl T4 Polynucleotide Kinase (T4 PNK, 10 U/μl), mix gently, and continue the incubation at 37°C for 1 h. Add 20 μl of water to the tube and
heat inactivate the enzyme at 65°C for 15 min. Centrifuge briefly and remove the unincorporated γ-32P ATP using a Quick Spin (Sephadex-G25, fine) column.

3. For use in PCR, dilute the labeled primer with water to a final volume of 80 μl (25 ng/μl concentration).

3. **Labeled Molecular Weight Marker DNA**

For determining the sizes of PCR products on a gel, use 32P-labeled molecular weight markers or sequencing ladders in one of the lanes. 1kbPlus ladder is labeled by exchange reaction of the enzyme T4 PNK. Other markers such as plasmid pBR 322 digested with Msp I (New England Biolabs) can be labeled in the same way.

1. In a 0.5 ml tube, add 2 μl (2 μg) of marker DNA and set up the reaction (25 μl) as follows:
   - Water 12 μl
   - 5 X exchange reaction buffer 5 μl
   - γ32P ATP (3000 Ci/mmol) 5 μl (50 μCi)

2. Prewarm the contents, initiate the reaction with 1 μl (10 units) of T4 PNK and continue incubation at 37°C for 30 min. Inactivate the enzyme and remove the unincorporated γ-32P ATP using a Quick Spin (Sephadex-G25 fine) column.

3. Take 1 μl of the 32P-labeled marker; add 5 μl of stop solution and 4 μl of water. Denature at 95°C for 3 min and chill the solution on ice. Load 2 μl of this sample as marker on the gel.

C. **cDNA SYNTHESIS**

1. Carry out cDNA synthesis using 6 μg of total RNA (10 μl volume) with 1 μl of T7-Sal-Oligo d(T)18-V (200 ng/μl stock) as the primer in a 0.5 ml tube. Denature the primer and RNA mix at 65°C for 5 min and anneal for 5 min on ice. Repeat this cycle once.

2. Aliquot 8 μl of first-strand reaction master mix to each tube and prewarm the tubes to 45°C.

   The first-strand reaction master mix (8 μl) will have the following components:
   - 5 X first strand buffer 4 μl
   - 0.1 M DTT 2 μl
   - 10 mM dNTPs 1 μl
   - RNase inhibitor (40 U/μl) 1 μl

3. Initiate the cDNA synthesis with 1 μl (200 units) of Superscript II Reverse transcriptase (final reaction is 20 μl) and continue the reaction for 1 h at 45°C. Use an air incubator or overlay the tubes with mineral oil to avoid evaporation during incubation.

4. Following first-strand synthesis, chill the tubes, centrifuge briefly, and add 130 μl of the second-strand reaction master mix to each tube.

   Master mix for second strand synthesis contains the following components:
   - Water 91 μl
   - 5 X second-strand buffer 30 μl
   - 10 mM dNTPs 3 μl
   - E. coli DNA polymerase (10 U/μl) 4 μl
   - E. coli DNA ligase (10 U/μl) 1 μl
   - RNase H (3 U/μl) 1 μl

   Carry out second strand cDNA synthesis in a final volume of 150 μl for 2 h at 16°C and stop the reaction with 10 μl of 0.5 M EDTA pH 8.0.

5. Clean up the cDNA by extracting once with phenol: chloroform (1:1 v/v) and once with water-saturated chloroform.

www.taq.ir
6. Precipitate the cDNA from the aqueous phase with 2.5 volumes of chilled ethanol in the presence of 1 µl (20 µg) of glycogen as carrier and 0.5 volumes of 7.5 M ammonium acetate.

7. Collect the cDNA precipitate by centrifugation at top speed for 15 min. Wash the pellet once with 70% ethanol.

8. Let the pellet dry at room temperature and dissolve it in 20 µl of water or 1 X TE buffer.

D. Restriction Enzyme Digestion of the cDNA

The cDNA thus prepared is digested separately with different restriction enzymes that have six nucleotide recognition sites. These enzymes are selected based on their frequency of cutting the cDNAs from a given species (human, rat, mouse, or plant source, etc.). The procedure below is to set up a cDNA digestion with Bgl II enzyme; however, we can use any restriction enzyme with the corresponding 10 X buffer to set up the digestion in the similar manner.

1. Prepare a restriction enzyme reaction mix as follows on ice.
   - Water: 33 µl
   - 10 X enzyme (Bgl II) buffer: 5 µl
   - Restriction enzyme (Bgl II): 2 µl (20 units)
   - Final volume: 40 µl

2. Take 2 µl of cDNA in a 0.5 ml tube and add 8 µl of the restriction enzyme reaction mix (final volume is 10 µl); carry out digestion for 2 h at 37°C and inactivate the enzyme for 20 min at 65°C.

E. Ligation of Digested cDNA with Fly Adapter

1. To a 0.5 ml tube on ice, add 88 µl of water and mix with 10 µl of T4 DNA ligase buffer (10 X, New England Biolabs) and 2 µl of T4 DNA ligase (400 U/µl). This is referred to as diluted ligase (8 U/µl).

2. In a 0.5 ml tube on ice take 20 µl of water, mix with 10 µl of T4 DNA ligase buffer (10 X), and add 10 µl of the diluted ligase (final volume is 40 µl and has 80 U Ligase). This is referred to as ligase reaction master mix.

3. Take a 0.5 ml tube on ice, add 2 µl of restriction enzyme (Bgl II) digested cDNA and mix with 1 µl (100 ng) of the annealed fly adapter with suitable overhang (5'-GATC overhang for Bgl II). Add 2 µl of ligase reaction master mix (4 U ligase/reaction and 1 X buffer concentration in the final reaction), mix the contents, and incubate at 16°C overnight.

F. PCR Amplification

Fly adapter–ligated cDNA template is PCR-amplified with N1N2-anchored oligo primer (3' primer) and 32P-labeled FA-1 (5' primer). Twelve separate PCR reactions are set up for each fly adapter–ligated cDNA, as there are 12 possible N1N2 anchors.

1. Aliquot 2 µl of a 2 µM stock solution of 3'-anchored primer [N1N2 anchored oligo d(T)18 with heel] to a PCR tube and mix with 18 µl of the PCR master mix to set up a 20 µl PCR reaction.

   The PCR master mix will have the following components per 18 µl.
   - Gene Amp 10 X PCR buffer (with MgCl2): 2 µl
   - 2 mM dNTPs: 2 µl
   - 32P-labeled primer FA-1: 2 µl
Restriction Endonucleolytic Analysis of Differentially Expressed Sequences: READS

Water 9 µl
AmpliTaq Gold (5 U/µl) 0.14 µl
Fly adapter–ligated cDNA template
Diluted 1:20 in water 2 µl
Adjust the final volume to 18 µl with water

2. Overlay the tubes with mineral oil and carry out READS-PCR as follows:
   12 min at 94°C
   5 cycles of 30 sec at 94°C, 1 min at 55°C, 2 min at 72°C
   25 cycles of 30 sec at 94°C, 2 min at 60°C, 1 min 72°C
   After the PCR store all samples at +4°C.

G. Sequencing Gel Analysis

1. Siliconize the smaller of the two sequencing gel plates with Sigmacote and clean the plates with ethanol. Assemble the glass plates and cast a 0.4 mm thick sequencing gel with gel-mix 6 acrylamide solution. (We used a comb that forms 36 wells of 6.8 mm width with a 2 mm space between each well.)
2. Assemble the gel in the electrophoresis unit and do a prerun for 1 h using 1 X Trisborate-EDTA (1 X TBE) buffer at 2000 V constant voltage.
3. Prepare the samples by heat denaturing 2 µl of the PCR products with 2 µl of the stop solution in a tube at 95°C for 3 min and chilling them quickly on ice. Disconnect power supply, clean the wells with 1 X TBE, and load the samples. Load 32P-labeled molecular weight markers in one of the lanes. Continue electrophoresis at 2000 V (constant) until the xylene cyanol dye has just come out of the gel.
4. Disconnect the power supply. Remove the siliconized small glass plate and transfer the gel onto a precut sheet of filter paper (Whatman 3 mm). Cover it with extra wide Saran wrap and dry the gel in a gel drier, with an additional layer of filter paper below to protect the gel from picking up any contaminants from the gel drier.

H. Autoradiography

Remove the dried gel from gel drier and stick a few pieces of Radtape (with details of the gel written on it) at four corners of the gel as orientation marks. Expose the gel to Biomax-MR x-ray film at room temperature overnight. Store the dried gels in zip-seal polyethylene bags for band recovery.

I. Recovery of cDNA Bands from the Gel and PCR Amplification

Based on the results from the autoradiogram, specific cDNA bands of interest can be recovered from the gel and PCR amplified for sequence analysis.

1. Mark next to the bands of interest on the autoradiogram with a fine point marker.
2. Mount the gel on a soft board covered with a couple of sheets of filter paper (Whatman 3 mm) below as padding. Place the autoradiogram on top of the gel, align its orientation marks to that of the dried gel, and hold the gel, and the x-ray film with thumb tacks.
3. Take a 21-gauge needle (1.5 in.), hold it vertically at the center of the band, and stab through the film into the gel. Apply just enough pressure so that the needle does not pass through the gel into the bottom padding. Carefully dip the end of the needle for a few seconds into a PCR tube containing 50 µl PCR reaction mix.
The PCR reaction mix has the following components:
- 10 X GeneAmp PCR buffer 5 µl
- 2 mM dNTPs 5 µl
- Primer FA-1 (25 ng/µl) 5 µl
- 3' anchored primer or heel primer (2 µM stock) 5 µl
- Water to adjust the final volume to 50 µl
- AmpliTaq Gold (5 U/µl) 0.14 µl

Prepare a master mix based on the number of bands to be amplified, aliquot 50 µl per tube, and then proceed with recovering the bands.

Overlay the sample with mineral oil and run a PCR reaction as follows:
- 12 min at 94°C
- 30 to 35 cycles of 30 sec at 94°C, 120 sec at 60°C, 60 sec at 72°C

Alternatively, a two-temperature PCR can also be used with the following conditions:
- 12 min at 94°C
- 30 to 35 cycles of 30 sec at 94°C, 150 sec at 68°C

Both conditions work well. If heel primers are used, anneal at 60°C followed by extension at 72°C.

J. PROCESSING OF THE PCR-AMPLIFIED SAMPLES FOR SEQUENCE ANALYSIS

Ensure that the PCR product is of the right size without any contaminating bands by analyzing an aliquot on a 2%-agarose gel (see Figure 28.2). If the product is pure enough and is of the expected size, remove the unincorporated primers and dephosphorylate the dNTPs using exonuclease I and shrimp alkaline phosphatase treatment, as follows:

1. EXO-SAP enzyme mix: mix shrimp alkaline phosphatase 1 U/µl and exonuclease I (10 U/µl) at a ratio of 10:1 v/v (1:1 U/U) in a 0.5 ml tube on ice. This mixture can be stored at −20°C.

2. Add 2 µl of EXO-SAP enzyme mix to 5 µl of the PCR product in a PCR tube and incubate the tubes at 37°C for 15 min and at 80°C for 15 min in a thermal cycler. Dilute the samples to approximately 10 ng DNA for every 100 bp in a final volume of 15 µl with water and submit the samples for automated sequencing with FA-1 as primer.

If the sample has other contaminating bands along with the correct size band, excise the specific DNA band from the gel and extract the DNA using a Qiagen gel extraction kit. This gel-extracted DNA can be either sequenced with FA-1 as primer or used for cloning into a plasmid vector.

Generally, direct sequencing of the PCR product results in good quality sequence. However, sometimes, when the sequence information is not clean enough for further bioinformatics analysis, it becomes essential to clone the fragments into a plasmid vector. There are several PCR cloning vector systems available commercially, such as PCR Script from Stratagene, TOPO cloning system from Invitrogen, and pGEM-T easy Vector system I from Promega Corporation. All three systems worked well in our hands using the suppliers’ protocols.

K. DATA DOCUMENTATION

The autoradiograms are scanned for analysis by using a UMAX -Mirage D-16L X-ray film scanner connected to a personal computer. Quantitative information on the relative band intensities from the dried sequencing gel can also be obtained by using a phosphorimager with Imagequant™ software (Molecular Dynamics). This information can be further analyzed with Spotfire Pro (www.spotfire.com) software for specific expression paradigms.
Restriction Endonucleolytic Analysis of Differentially Expressed Sequences: READS

**L. TECHNICAL COMMENTS**

READS is a very reproducible, consistent approach to study differential gene expression and is very easy to carry out. The success of this analysis depends on RNA that is of good quality and free of contaminating genomic DNA.

The current protocol incorporated several improvements and modifications. Prashar and Weissman used an N1N2 (dinucleotide) anchored oligo d(T) as primer to make first-strand cDNA. Since a dinucleotide anchor results in 12 possible primers, their method involved 12 individual cDNA preparations. By contrast, we used a mixture of N1 anchored oligo d(T)₁₈ (oligo d(T)₁₈₋₋ where V is A, C, or G) in a single cDNA preparation per RNA. Therefore, only 1/12 of the RNA (e.g., 6 μg or less) is needed by our approach. The cDNA thus prepared is digested with a restriction enzyme, ligated with the βγ adapter, and subjected to PCR amplification. At the PCR stage, we used each of the N1N2 anchored primers to set up 12 individual PCR reactions. These changes not only enabled us to carry out READS using small amounts of each RNA but also avoided the need to make multiple cDNA preparations from every RNA. cDNA made from 6 μg of total RNA is sufficient for generating the display patterns for about 10 restriction enzymes using all the anchored primers. It is also possible to use much less RNA (1 to 6 μg), and a method has been used to make cDNA even with submicrogram quantities of RNA for subsequent (micro) READS analysis.

One could design the fly adapters suitable for other restriction enzymes by changing the four overhang bases on FA-2. For 3’ overhang enzymes (such as Pst I), FA-1 carries the four-base overhang. In both cases ³²P labeled FA-1 is used for PCR. We also used γ³³P-ATP for primer labeling. This resulted in sharper bands. Also ³³P is a low energy isotope that may be safer than ³²P isotope. As some restriction enzymes are not effectively heat inactivated, design the overhangs so that the restriction enzyme site is not regenerated once the βγ adapter is ligated to the cDNA.

We selected a two-stage PCR procedure to amplify the fly adapter–ligated cDNA. In the first stage, five cycles of low temperature annealing at 55°C help efficient annealing of the template and the anchored primers. In the second stage, the annealing temperature is raised to 60°C to

**FIGURE 28.2** Agarose gel analysis of the PCR amplified cDNA bands. Differentially regulated cDNA bands isolated from a READSTM gel were PCR amplified as described in methods. An aliquot of the PCR product was analyzed on a 2% agarose gel (Lanes 1 to 7). Lane M shows a 100-bp ladder marker.
maintain the specificity. We recommend trying different dilutions of the ligated cDNA template for the READS-PCR reaction to find out the optimal dilution and quantity. A good starting point is to use 2 μl of 1:20 diluted template for PCR reaction. Too much template results in smears patterns, while too little template results in faint and/or missing bands. Certain anchored primers (e.g., GG, GC, CG, and CC anchors) incorporate more radioactive primer and hence result in darker bands as compared to certain other primers (e.g., GT, AT, CT, and GA). The PCR conditions described above work well for all the anchored primers; however, it is also possible to employ higher annealing temperature or reduce the number of cycles with some of the anchored primers (as in the case of GG, GC, CG, and CC), and the annealing temperature could be lowered or the number of cycles increased for certain other primers (GT, AT, CT, and GA).

Direct sequencing of the PCR product results in good quality sequence, and we always prefer to use this approach. We clone the product for sequencing only when a given PCR product results in an unreadable sequence (this could be due to the presence of multiple cDNAs with same size). While this results in a high quality sequence, we need to sequence the DNA from statistically significant number of colonies to ensure that a given sequence is representative so that it can be attributed to a specific cDNA band on the gel with confidence. If the sequence obtained corresponds to a sequence from public databases, we can analyze the sequence for the N1N2 and the restriction enzyme site that were used in the study. For a cDNA with known sequence information, we can predict and confirm the exact position on the gel.

IV. DISCUSSION

As shown in Figure 28.3, neutrophils exhibit distinct changes in mRNA expression in response to LPS treatment. A part of the gel picture from Figure 28.3 is enlarged in Figure 28.4 to show these changes.
Restriction Endonucleolytic Analysis of Differentially Expressed Sequences: READS

changes more clearly. By collecting samples at different time points, we were able to study the kinetics of this induction. In our laboratory we used READS approach to study the mRNA expression changes in neutrophils in response to bacteria and other agents such as LPS.

Total reagent cost to analyze one RNA sample by READS analysis with one restriction enzyme is about $60 and this does not include the costs involved in sequencing the differentially regulated cDNA bands recovered in the study. A given study may involve a set of RNA samples analyzed in parallel to look for the differential regulation. Sequencing costs for band identification usually occur only once per set of samples as bands are readily comparable across gels from similar specimens. Although preparation and running of READS display gels require somewhat more skill than does array hybridization, the total costs of the procedure compare favorably with the use of commercial oligonucleotide chips. For smaller laboratories or a limited number of analyses, the effort and investment in equipment is much less for READS than that for cDNA arrays, and there is less chance of confusion of DNA fragments.

The method of selective amplification of DNA fragments by use of Y-shaped adapters has proven highly selective in our hands and is more general than simple gel-based cDNA displays. For example, the procedure could be applied to selectively amplify and display subsets of cDNA fragments previously selected to contain internal mismatched nucleotides, consequent upon genetic polymorphisms, or to identify DNA fragments ligated together as a result of three-dimensional proximity following in vivo cleavage and religation by the procedure of Dekker et al.

While the gel display–based approaches are rather labor intensive, in principle this could be expedited by developing reproducible approaches for display of amplified fragments by capillary electrophoresis by which multiple runs could be examined in parallel in commercial DNA sequencing instruments.

REFERENCES
1. READS™ is a trade marked technology of Gene Logic Inc., and is covered by U.S. patent no. 5712126.
20. *SuperScript Choice system for cDNA synthesis*, Instruction manual from GIBCO-BRL.
I. INTRODUCTION

The identification of differentially expressed genes can provide insights into the molecular mechanisms underlying disease progression or biological development. To facilitate the discovery of differentially expressed genes, a variety of methods have been developed in recent years including Differential Display PCR, RNA fingerprinting, SAGE, Real-time Quantitative PCR (TaqMan), subtractive hybridization, and hybridization to gene arrays of various formats. Among these methods, subtractive hybridization is unique for being able to preferentially enrich differentially expressed genes. The method was first used by Bautz and Reilly to purify phage T4 mRNA in the mid-1960s. Pure subtractive methodologies are of limited use because they require large quantities of mRNA to drive hybridization to completion. Moreover, subsequent cloning is difficult because only a tiny amount of cDNA remains after hybridization. The method was improved when Duguid and Dinauer added generic linkers to the cDNA allowing PCR amplification of tester cDNA between hybridization cycles. Advancing further, Lisitsyn et al. introduced a method for selective PCR amplification of double-strand tester DNA formed after subtractive hybridization, eliminating the need to physically separate it from the other components. Subsequently, Siebert et al. invented Suppression PCR, in which DNA with long and complementary 3' and 5' ends was “suppressed” during PCR amplification due to the formation of a “panhandle,” a secondary structure which prohibits PCR primer binding. Diatchenko et al. integrated all aforementioned technical innovations into Suppression Subtractive Hybridization PCR (SSH PCR), targeting the subtracted and
normalized tester DNA for selective amplification, resulting in over 1000-fold enrichment of differentially expressed gene in a single round of hybridization.\textsuperscript{5} The methodology of SSH PCR has been fully described by its inventors previously.\textsuperscript{5,12,13} Here, we will concentrate on its key technical innovations and demonstrate its efficacy in isolating differentially expressed genes.

## II. MATERIALS AND METHODS

### A. PRIMERS

The primer sequences listed in Table 29.1 are described in the Clontech PCR-Select cDNA Subtraction Kit (Clontech, Palo Alto, CA, www.clontech.com), but can be synthesized by various vendors. The primers were resuspended in double-distilled water to obtain a stock solution of 1.0 mM.

### B. PREPARATION OF cDNA

In general, SSH PCR requires more than 0.5 μg of cDNA for subtractive enrichment of differentially expressed genes, which, in many practical applications, is difficult to meet. We employed the SMART PCR amplification method described recently by Zhumabayeva et al. to generate microgram range of cDNA from as little as 50 ng of total RNA from primary human fibroblast cell cultures.\textsuperscript{14} The same batch of SMART PCR cDNA was used for both tester and driver to insure the uniformity of background. The amplified cDNA was then digested with the restriction enzyme Rsa I and purified by ethanol precipitation. Rsa I is a 4-base restriction enzyme that generates DNA fragments of suitable sizes for PCR amplification. Digested cDNA was quantified by UV spectrum at 260 nm and used as driver or tester. In our experiments, \textit{Hae III} digested phage\textit{φX174} was added to driver and tester at this stage to simulate differentially expressed genes.

### C. LIGATION OF ADAPTERS TO TESTER cDNA

Each ligation reaction consisted of 50 mM Tris-HCl, pH 7.8, 10 mM MgCl\textsubscript{2}, 2.0 mM DTT, 300 μM ATP, 400 U of T4 DNA ligase, 2.0 μM Adapter 1 or 2, 120 ng of cDNA, and H\textsubscript{2}O in a volume of 10 μl. The reaction was incubated at 16°C for 16 h. It was stopped by 1.0 μl of 200 mM EDTA and heated at 72°C for 5 min.

### D. FIRST HYBRIDIZATION

The first hybridization reaction consisted of 1 X Hybridization Buffer (50 mM HEPES-HCl, pH 7.5, 500 mM NaCl, 0.2 mM EDTA), 18 ng of tester with Adapter 1 or Adapter 2, 540 ng of driver, and H\textsubscript{2}O in a volume of 4 μl. The Adapter 1 and 2 reactions were carried out in different tubes. They were covered with a layer of paraffin oil, denatured at 98°C for 90 sec, and renatured at 68°C for 8 h.

### TABLE 29.1

<table>
<thead>
<tr>
<th>Name</th>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adapter 1</td>
<td>5'-CTAATACGACTCACTATAGGGTGACTCAGCAGG-3'</td>
</tr>
<tr>
<td></td>
<td>3'-GGCCCGTCCA-5'</td>
</tr>
<tr>
<td>Adapter 2</td>
<td>5'-CTAATACGACTCACTATAGGGTGACTGAGCAG-3'</td>
</tr>
<tr>
<td></td>
<td>3'-GCCCGTCCA-5'</td>
</tr>
<tr>
<td>PCR Primer 1</td>
<td>5'-CTAATACGACTCACTATAGGGTGACTGAGCAG-3'</td>
</tr>
<tr>
<td>Nested PCR Primer 1</td>
<td>5'-TCAGGCGGCGGCGGCGGCGGAGGT-3'</td>
</tr>
<tr>
<td>Nested PCR Primer 2</td>
<td>5'-AGCGTGGTGCGGCGGCGGAGGT-3'</td>
</tr>
</tbody>
</table>

www.taq.ir
E. **SECOND HYBRIDIZATION**

The 360 ng of driver was denatured in 1 X Hybridization Buffer in a volume of 4 µl. The reaction was covered with a layer of paraffin oil heated at 98°C for 1.5 min and incubated at 68°C. The first hybridization of Adapter 1 and 2 were added sequentially to the denatured driver and the whole mixture was incubated at 68°C for 20 h. The second hybridization was diluted with 200 µl of dilution buffer (20 mM HEPES-HCl, pH 8.3, 50 mM NaCl, 0.2 mM EDTA heated at 68°C for 7 min and stored at –20°C.

F. **PCR AMPLIFICATION**

Extension of 3’ ends and PCR amplification were carried out in the same tube by an MJ Research PT-200 thermocycler. The reaction consisted of 1 X PCR buffer [40 mM Tricine-KOH, pH 8.7, 15 mM KOAc, 3.5 mM Mg(OAc)$_2$, 3.75 µg/ml BSA, 0.005% Tween 20, 0.005% Nonidet-P40], 200 µM dNTP, 400 nM PCR Primer 1, 0.5 µl of 50 X Advantage 2 Polymerase Mix (Clontech, Palo Alto, CA), 1.0 µl of the diluted second hybridization, and H$_2$O in a volume of 25 µl. It was incubated at 75°C for 5 min to extend 3’ end, followed immediately by PCR amplification: 95°C for 25 sec, 27 cycles of 94°C for 10 sec, 66°C for 30 sec, and 72°C for 90 sec. The PCR mixture was diluted 10 times and amplified again by nested PCR. The nested PCR reaction consisted of 1 X PCR buffer, 200 µM dNTP, 400 nM Nested PCR Primer 1, 400 nM Nested PCR Primer 2, 0.5 µl of 50 X Advantage 2 Polymerase Mix, 1.0 µl of the diluted primary PCR products, and H$_2$O in a volume of 25 µl. It was amplified by 12 cycles, consisting of 94°C for 10 sec, 68°C for 30 sec, and 72°C for 90 sec. The PCR products were separated by electrophoresis on 2% agarose gel, stained in 500 ng/ml ethidium bromide, and examined under UV light at 254 nm.

III. **RESULTS**

A. **OUTLINE OF SSH PCR**

Two outstanding features of SSH PCR deserve elaboration (Figure 29.1): (1) it preferentially amplifies subtracted cDNA without physical separation from other components, (2) it preferentially amplifies the normalized tester cDNA without physical separation from other components. The first innovation was introduced by Lisitsyn for Representational Differential Analysis (RDA), another subtractive DNA technology. The key is to ligate dephosphorylated adapters to only the 5’ ends of tester cDNA. Thus, following subtractive hybridization and 3’ end extension, only tester/tester double-stranded cDNA (ds cDNA) can have both 5’ and 3’ ends fitted with the adapter and its complementary sequences and thereby be amplified by subsequent PCR. The second innovation combines the second-order kinetics of DNA hybridization and Suppression PCR. As shown in Figure 29.1, tester cDNA is normalized in single-stranded cDNA (ss cDNA) during the first hybridization. At this stage, ds cDNA cannot be amplified by PCR since the specially designed long adapters suppress PCR primer binding. Only in the second hybridization, when the normalized ss cDNA from the first hybridizations are combined and renatured, certain portions of ds cDNA will be fitted with heterologous adapters on both ends. Only these will be amplified exponentially by PCR. In both hybridizations, denatured driver is added to subtract tester cDNA to reduce commonly expressed genes.

B. **EFFICACY OF SSH PCR**

The efficacy of SSH PCR was measured by artificially adding φX174 Hae III digest to tester and driver to simulate differentially expressed genes and enriching it by SSH PCR. The intensity of the φX174 bands was compared with background, providing an estimate of efficacy of SSH PCR. First, the efficacy of enriching “all or nothing” differentially expressed genes was tested. In the experiment, φX174 was present in the tester at different concentrations while absent in the driver.
The results are demonstrated in Figure 29.2. It shows that \( \phi x174 \) fragments could be enriched to distinct bands when its fractional concentration was more than 0.01% in tester (lane 2, 3). Similar intensity of \( \phi x174 \) Hae III bands in lane 2 and 3 suggests a comparable enrichment for 1.0% and 0.1% \( \phi x174 \). By contrast, the absence of a distinct band of \( \phi x174 \) in lane 4, 5, and 6 suggests that the predominant SSH PCR products were not “differentially expressed” \( \phi x174 \) fragments but randomly amplified fibroblast cDNAs.

Next, the efficacy of enriching ratio differentially expressed genes was tested. In the experiments, \( \phi x174 \) was added to both tester and driver, but at different quantities. In the first series, the tester contained \( \phi x174 \) at a fixed amount of 1.0%, while various amounts ranging from \( 10^{-3} \) to 1% were added to driver. The results are shown in Figure 29.3. Distinct, enriched bands were visible only when “differentially expressed” \( \phi x174 \) DNA was five-fold or more concentrated in tester than that in driver (lane 4, 5, 6, and 7). When the concentration ratio was less than 5, no distinguishable \( \phi x174 \) DNA bands could be seen (lane 2 and 3), suggesting that predominant SSH PCR products were randomly amplified fibroblast cDNAs.

To further examine the effect of concentration ratios on the efficacy of SSH PCR, we made another series of tester and driver. In this series, tester contained a fixed amount of \( \phi x174 \), at 0.1% while various amounts of \( \phi x174 \) DNA were added to the driver. Thus, the absolute amount of \( \phi x174 \) DNA was 1/10 of that in the previous experiment, while the corresponding concentration ratios remained the same. The results, shown in Figure 29.4, are almost identical to those in Figure 29.3. They demonstrate that effective enrichment by SSH PCR depends mainly on

FIGURE 29.1 Outline of SSH PCR. The procedure consists of two subtractive hybridizations followed by PCR amplifications. While both hybridizations are designed to enrich tester-specific cDNAs, the first hybridization also serves to normalize the cDNA population. The tester cDNAs are fitted with two different long adapters at their 5’ ends. These form a “panhandle” structure which suppresses PCR amplification of unwanted cDNAs.
FIGURE 29.2 Enrichment of “all or nothing” differentially expressed genes by SSH PCR. As indicated in the figure, $\phi x 174$ Hae III of various amounts of DNA was added in tester, while none was added in driver. SSH PCR was carried out as described in the section Materials and Methods. Lane 1, $\phi x 174$ Hae III DNA marker; lane 2, 1.0% of $\phi x 174$ Hae III in tester; lane 3, 0.1% of $\phi x 174$ Hae III in tester; lane 4, 0.01% of $\phi x 174$ Hae III in tester; lane 5, 0.001% of $\phi x 174$ Hae III in tester; lane 6, 0.0001% of $\phi x 174$ Hae III in tester.

FIGURE 29.3 Enrichment of ratio differentially expressed genes by SSH PCR. A fixed amount of 1.0% of $\phi x 174$ Hae III was added in tester, while various amounts of $\phi x 174$ Hae III was added in driver. SSH PCR was carried out as described in the section Materials and Methods. Lane 1, $\phi x 174$ Hae III marker; lane 2, 1.0% of $\phi x 174$ Hae III in driver; lane 3, 0.2% of $\phi x 174$ Hae III in driver; lane 4, 0.04% of $\phi x 174$ Hae III in driver; lane 5, 0.008% of $\phi x 174$ Hae III in driver; lane 6, 0.0016% of $\phi x 174$ Hae III in driver; lane 7, none of $\phi x 174$ Hae III in driver.
the concentration ratio of the target gene(s) between two cDNA preparations. It is far more effective for genes that show large differences in their expression levels.

IV. DISCUSSIONS

A. THEORETICAL CONSIDERATIONS

The notion that the efficacy of SSH PCR is influenced by the concentration ratios of expressed genes in tester and driver is confirmed by a theoretical model. Assuming that DNA hybridization obeys the ideal second-order kinetics, the single-strand cDNA concentration after the first round of hybridization can be described by Equation 1:

$$C_t = \frac{C_0}{C_0kt + 1}$$  \hspace{1cm} (1)

where $C_0$ is the starting ss cDNA concentration and $C_t$ is the ss cDNA concentration at time $t$. When $C_0kt >> 1$, Equation 1 simplifies to Equation 2:

$$C_t = \frac{1}{kt}$$  \hspace{1cm} (2)

which implies that when hybridization time is long enough, the ssDNA concentration is determined mainly by its hybridization rate constant $k$ and hybridization time $t$, independent of its starting concentration $C_0$, which forms the basis of normalization in the first hybridization reaction.

Because the ss cDNAs consist of both tester cDNAs which have adapter, and driver cDNAs which do not, the concentration of ss cDNA with adapters can be calculated by Equation 3:

FIGURE 29.4 Enrichment of ratio differentially expressed genes by SSH PCR. Fixed amount of 0.1% of φx174 Hae III was added in tester, while various amounts of φx174 Hae III was added in driver. SSH PCR was carried out as described in the section Materials and Methods. Lane 1, φx174 Hae III DNA marker; lane 2, 0.1% of φx174 Hae III in driver, lane 3, 0.02% of φx174 Hae III in driver; lane 4, 0.004% of φx174 Hae III in driver; lane 5, 0.0008% of φx174 Hae III in driver; lane 6, 0.00016% of φx174 Hae III in driver; lane 7, none of φx174 Hae III in driver.
Suppression Subtractive Hybridization PCR

\[ C_t = \frac{C_t}{1 + N / R} \]  

(3)

where \( C_t \) is the concentration of a target single-stranded cDNA with adapter, \( N \) is the ratio of the driver to tester in the first hybridization, and the \( R \) is the concentration ratio of the target cDNA in tester to that in driver.

In the second hybridization, ss cDNAs from the first hybridization are mixed with new denatured driver cDNAs. The reaction is carried out over a longer time period to drive hybridization to completion. This reaction can be represented by the formula:

\[ (A + A' + A'')(B + B' + B') \]

where \( A \) and \( B \) are a single-stranded cDNA and its complementary strand. \( A' \) and \( B' \) are fitted with Adapter 1, and \( A'' \) and \( B'' \) are fitted with Adapter 2. Following the second hybridization, only the double-strand cDNAs with two different adapters at each end (\( A'B'' \) and \( A'B' \)) can be amplified by PCR. The amount of \( A'B'' + A'B' \) can be calculated by Equation 4:

\[ A'B'' + A''B' = \frac{A'B'' + A''B'}{(A + A' + A'')(B + B' + B'')} (A + A' + A'') \]  

(4)

Given that \( A = B = MC_0 / R \), where \( M \) is the ratio of driver to tester in the second hybridization and \( R \) is the concentration ratio of a target cDNA of tester to driver, and \( A' = B' = A'' = B'' = C' = C_t / (1 + N / R) \) from Equation 3, the ds cDNA with hetero-adapters can be calculated by Equation 5:

\[ A'B' + A'A' B' = \frac{2C_t^2}{(1 + N / R)(MC_0 / R + NMC_0 / R^2 + 2C_t)} \]  

(5)

where \( C_t \) is the concentration of remaining ss cDNA after the first hybridization, \( N \) is the ratio of driver to tester in the first hybridization (30 in our experiments), \( M \) is the ratio of driver to tester in the second hybridization (5 in our experiments), and the \( R \) is the concentration ratio of the target cDNA in tester to that in driver.

We can make two approximations: (1) we ignore the relatively small amount of linear amplification of cDNAs fitted with only one adapter; (2) we assume that all cDNAs fitted with two adapters are exponentially amplified with equal efficiency. Then, Equation 6 gives the relative amount of all SSH PCR products, which leads to the following conclusions. First, where target cDNA is present only in tester and not in driver, \( R = \infty \), and \( A'B'' + A'B' = C_t = 1 / kr \) (Equation 2). This means that every “all or nothing” differentially expressed cDNA will be enriched to a fixed level regardless of its starting concentration. Second, where target cDNA is present in both tester and driver at slightly different concentrations (e.g., \( R < 10 \)), Equation 5 can be simplified to:

\[ A'B'' + A'B' A'' = \frac{2R^2C_t^2}{N^2MC_0} \]  

(6)

This predicts that the enrichment of a differentially expressed gene is proportional to the cube of \( R \), favoring genes with large differences in expression.

As demonstrated in Figure 29.2, Figure 29.3, and Figure 29.4, the theoretical predictions have been largely confirmed. However, the model fails to predict the inability of SSH PCR to enrich targets of low concentrations (Figure 29.1, lane 4, 5, and 6). It might be difficult to drive the second
hybridization to completion when the concentration of the target cDNA is too low, for example, less than 0.01%. Because formation of ds cDNA is required for 3’ extension and thereby PCR amplification, rare targets will be underrepresented even if they are differentially expressed. Consequently, primary applications of SSH PCR should be those where dramatic alteration of gene expression can be expected, such as viral infection or comparisons between different tissues. SSH PCR would be ineffective in finding the targets of small incremental changes, such as diseased vs. normal tissues over a time course.

Presence of many genes with no differences in their expression complicates SSH analysis. Often, a bacterial library of SSH PCR cDNA is constructed and probed by both tester and driver, and only the genes that display differential hybridization will be chosen for further studies. Hybridization methods, such as microarrays, are powerful tools to discover differentially expressed genes so long as the cDNA library is large enough. However, using SSH PCR to enrich differentially expressed genes increases the chance of finding differentially expressed genes, even with a small library.

**B. Troubleshooting**

We generally found SSH PCR to be robust and reliable. However, the success of SSH PCR is critically influenced by the efficiency of the restriction enzyme digestion and the adapter ligation. It is advised that both steps be monitored closely during the procedure. For enzymatic digestion, we ran aliquots of cDNA on agarose electrophoresis gel before and after enzymatic digestion. Complete enzymatic digestion is indicated by downshift of cDNA molecular weight from 0.5–10 kb of cDNA to 0.1–3 kb of digested cDNA. For adapter ligation, we compared PCR amplification efficiency between two PCR reactions — one with two internal control 3’ and 5’ primers, and the other with the internal control 3’ primer and an adapter-derived primer. To design the PCR primers we first chose a target molecule with known cDNA sequence, such as cattle GAPDH, and then designed 3’ and 5’ PCR primers inside the Rsa I restriction fragment. The PCR product with the adapter primer should be at least 25% of the one with two internal control primers; otherwise ligation efficiency is deemed to be too low. As precaution, it is always a good practice to run a control SSH PCR with added DNA to simulate differentially expressed genes, such as Hae III digested φx174, to check overall SSH PCR efficiency.

**REFERENCES**


Section III

Preparative Application
Amplification of Unknown Sequences
30 PCR-Based Techniques for Cloning an Unknown DNA Fragment Adjacent to a Known Integrant

Eric Ka-Wai Hui and Szecheng J. Lo

CONTENTS

I. Introduction.................................................................................................................313
II. Strategies..................................................................................................................313
   A. PCR After Intramolecular Circularization.............................................................313
   B. Interspersed Repetitive Sequence (IRS)-PCR-Based Methods..........................314
   C. Ligation Anchored (LA)-PCR-Based Methods....................................................316
   D. Arbitrarily Primed (AP)-PCR-Based Method......................................................316
III. Conclusion...............................................................................................................319
Acknowledgments...........................................................................................................321
References......................................................................................................................321

I. INTRODUCTION

To understand the mechanism and biological effect of viral or foreign DNA insertions, characterization of DNA sequences which contain the integrant-host junction (IHJ) becomes important. The PCR-based strategy to amplify unknown DNA sequences that flank a known integrant has the advantage of simplicity and alacrity. In general, PCR requires a pair of primers both of which are constructed from a known sequence. This is, however, not applicable to study the integration site of a virus or transposable element since only the DNA sequence of the integrant will be known in advance. To circumvent this problem, a variety of PCR-based techniques have been developed. We attempt in this chapter to summarize these methods and discuss the limitation of each.

II. STRATEGIES

A. PCR AFTER INTRAMOLECULAR CIRCULARIZATION

The first method to clone an unknown DNA sequence flanking a known integrant, inverse PCR (iPCR), was developed in 1980s.1,2 The basic principle of iPCR is shown in Figure 30.1. After the digestion of genomic DNA with a restriction enzyme (step 1), intramolecular self-ligation of these DNA fragments (step 2) generate a circularized form of DNA. Within the circularized form of DNA, a pair of primers perfectly annealing to the known integrant is applied to amplify the unknown DNA fragment by a conventional PCR method (step 3). The intramolecular circularization of the template genomic
PCR Technology: Current Innovations, Second Edition

DNA, hence, is the key step of this technique. In general, a 5-kb DNA fragment can be obtained from iPCR; to obtain a shorter (1 kb) or a longer (10 kb) length of DNA, several modifications such as partial inverse PCR (PI-PCR), long-distance iPCR, long-range iPCR (LR-iPCR), and long inverse PCR (LI-iPCR) have been developed and reviewed.\(^3\)–\(^5\) Recently, iPCR using a retroviral tag has been applied to identify new cancer genes and potential disease loci in mouse leukemia.\(^6\),\(^7\)

The major advantage of iPCR is in using two known specific primers to amplify the unknown sequence. However, it requires fastidious optimization of the intramolecular DNA ligation step or will result in no iPCR products. Furthermore, noncircularized DNA fragments, intermolecular ligation products, and free integrants may interfere with the iPCR or give a false-positive result.

B. INTERSPERSED REPETITIVE SEQUENCE (IRS)-PCR-BASED METHODS

Alternatively, one may design the second primer such that it anneals to an interspersed repetitive sequence (IRS). IRS elements, such as Alu repeats, are present at intervals of about 3 to 6 kb in the human genome.\(^8\) Alu element-mediated PCR (Alu-PCR) was originally developed to identify human genomes within a nonhuman genomic background.\(^9\) “Novel Alu-PCR,”\(^10\) a later modification, applies two specific primers — one ISP to the known integrant sequence and the other primer specific on the Alu sequence — to amplify the virus integration (Figure 30.2).

Obviously, a major danger with this method is the generation of false-positive PCR products from an inter-Alu (Alu–Alu) amplification. Therefore, two technical tricks have been applied.\(^10\),\(^11\) First, the ISP amounts in the PCR reaction are at least 10-fold higher than the Alu-specific primers (Figure 30.2, step 1). This “asymmetric amplification” does not favor Alu–Alu amplification. Secondly, the primers are synthesized by deoxuryridine triphosphates (dUTPs) instead of deoxythymidine triphosphates (dTTPs), which can then be digested by uracil DNA glycosylase (UDG) after

---

**FIGURE 30.1** Schematic diagram showing the principle of iPCR. Two complementary strands of genomic DNA are shown at the top. The heavy and thin line regions depict the integrant fragment (known region) and genomic DNA (unknown region), respectively. The positions of both left and right integrant-host junctions (IHJs) are indicated as L and R in hexagon boxes, respectively. Integrant specific primers (ISP) are shown as black arrowheads. The cross marks on thick arrows indicate that no primer annealing will occur in subsequent cycle and no further amplification products will be made.
PCR-Based Techniques for Cloning an Unknown DNA Fragment

The first 10 to 15 cycles (Figure 30.2, step 2). The treatment of UDG can efficiently destroy the Alu–Alu amplified DNA fragments.

For species lacking Alu elements, other IRS have been used under the same approach. Examples are LINE-PCR\(^{12}\) and B1-PCR,\(^ {13}\) using long interspersed nuclear element (LINE) and mouse specific IRS B1 family sequences, respectively.

IRS-based PCR seems to have several advantages over iPCR. First of all, a smaller amount of DNA is needed. Second, an intramolecular ligation reaction is not required. Third, IRS-PCR avoids
possible false-positive results from a contaminating episome. However, IRS-PCR is not well applicable on a short distance between the integrant and IRS element. Unfortunately, many viruses including hepatitis B virus (HBV) tend to insert adjacent to or into the repetitive sequence.\textsuperscript{14}

\section*{C. Ligation Anchored (LA)-PCR-Based Methods}

LA-PCR links an oligonucleotide cassette unit, called adapter or linker, to the genome fragment under investigation. Therefore, these methods require one ISP, one cassette unit-specific primer, and a suitable genomic restriction site downstream from the ISP.\textsuperscript{15}

Vectorette-PCR begins with a genomic DNA digestion (Figure 30.3A, step 1) and then ligation with a vectorette unit (Figure 30.3A, step 2). The flanking sequences are subsequently amplified by using an ISP and vectorette-specific primer. The vectorette unit is bubble-shaped, and contains a central noncomplementarily mismatched region. This structure ensures that exponential amplification occurs only after the second round of primer extension (Figure 30.3A, step 3).

Vectorette-PCR has been modified to a “multistep-touchdown vectorette” (MTV)-PCR\textsuperscript{16} and “long-distance vectorette” (LDV)-PCR\textsuperscript{17} which are suitable for analysis of fragments with a high CG content or large size (up to 5 kb), respectively.

However, end-repair priming on a free cohesive end of both the unligated free vectorette unit and an unknown genomic region might create undesirable amplification products.\textsuperscript{4,18,19} End-repair priming occurs in the first cycle of PCR, filling the ends which are then able to anneal together after denaturation and renaturation steps (Figure 30.3B). To eliminate the end-repair priming effect, the “splinkerette” (splinkerette-PCR) was therefore designed as a hairpin structure rather than a central mismatch (inset of Figure 30.3B).\textsuperscript{18}

Other methods for IHJ cloning, such as capture PCR combined with vectorette-PCR and panhandle PCR, have also been developed. Capture PCR employs biotinylated primers for the extension reactions, and the amplified products are then enriched via streptavidin-linked beads. By using this combination method, Laging et al. were able to isolate segments of homologous genes with only one conserved amino acid.\textsuperscript{20} Panhandle PCR requires a single-strand adapter and is, in principle, a combination of LA-PCR and iPCR. Because the adapter contains an inverted sequence of the integrant, intrastrand annealing can occur after denaturation and a panhandle structure is formed when the extension of recessed 3’ end is filled.\textsuperscript{21}

\section*{D. Arbitrarily Primed (AP)-PCR-Based Method}

Basically, arbitrarily primed PCR (AP-PCR) is using nonspecific “arbitrary degenerate primers” (ADP) for PCR amplification. However, due to the nonspecific annealing of the ADP to a template DNA, it can create a lot of nonspecific PCR products. To overcome this problem, scientists have designed a consecutive PCR reaction with nested sequence-specific primers and a shorter ADP called thermal asymmetric interlaced (TAIL)-PCR.\textsuperscript{22,23}

The principle of TAIL-PCR is to control the amplification of specific vs. nonspecific products by varying the annealing temperature during the amplification reaction. First, total genomic DNA is amplified using the ADP together with a specific primer (ISP1) with a relatively high $T_m$ in a series of reaction cycles with different annealing temperatures (Figure 30.4, step 1). During the high stringency cycle (Figure 30.4, left box on step 1) only the long specific primer ISP1 can efficiently anneal to DNA template. Therefore, product I is amplified while little or no amplification occurs for nonspecific products II which are primed at both ends by the ADP. In the following low stringency cycle (Figure 30.4, second left box on step 1), both ISP1 and ADP can anneal to the template. By repeating this process (TAIL-cycling; Figure 30.4, right box on step 1), it is possible to amplify specific product I preferentially over nontarget products II. In a subsequent secondary reaction of TAIL-cycling (Figure 30.4, step 2) and tertiary reaction of low stringency (Figure 30.4, step 3), product I would be the preferential products in lower
FIGURE 30.3 (A) Schematic diagram showing the principle of vectorette-PCR. Two complementary strands of genomic DNA are shown at the top. The black heavy line region and thin line regions represent the integrant fragment (known region) and genomic DNA (unknown region), respectively. The positions of both left and right IHJs are indicated as L and R in hexagon boxes, respectively. The vectorette unit has been shown. Integrant specific primers for right IHJ (ISPr), and vectorette specific primer are shown as black and white arrowheads, respectively. The cross marks on thick arrows indicate that no primer annealing will occur in subsequent cycle and no further amplification products will make. (B) Schematic diagram showing the effect of “end-repair priming." A nonspecific PCR product is amplified by using a pair of vectorette primers. The inset shows a splinkerette primer and the hairpin structure formed after the first cycle of PCR.
FIGURE 30.4 Schematic diagram showing the principle of TAIL-PCR. Two complementary strands of genomic DNA are shown at the top. The heavy and thin line regions represent the integrant fragment (known region) and genomic DNA (unknown region), respectively. The positions of both left and right IHJs are indicated as L and R in hexagon boxes, respectively. Integrant specific primers (ISP1, ISP2, and ISP3) and arbitrary degenerate primer (ADP) are shown as black and white arrowheads. The PCR program with different stringencies and cycle number is shown on the top of the box. Letters D, A, and P on the bottom of the box represent denature, annealing, and polymerization step in PCR cycle, respectively.
PCR-Based Techniques for Cloning an Unknown DNA Fragment

background. Although Product III, primed at both ends by the long ISP1, can also arise through mispriming in the primary reaction (Figure 30.4, step 1), such undesired products are diluted out in subsequent secondary and tertiary reactions.

TAIL-PCR does not require the digestion of genomic DNA or ligation of DNA unit prior to amplification, which is one of its advantages. TAIL-PCR, in addition, requires a very low template DNA quantity (~1 to 10 ng). By contrast, the ligation-dependent PCR methods such as iPCR and LA-PCR require much more DNA (~1 to 10 µg). However, the drawback of TAIL-PCR is a high background of nonspecific products.

III. CONCLUSION

PCR-based techniques for cloning an unknown DNA have several advantages over the traditional method of genomic library construction including simplicity, alacrity, high specificity, and

<table>
<thead>
<tr>
<th>Virus</th>
<th>Methods</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hepatitis B virus (HBV)</td>
<td>iPCR</td>
<td>a</td>
</tr>
<tr>
<td>Human immunodeficiency virus type-1 (HIV-1)</td>
<td>iPCR</td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>Alu-PCR</td>
<td>c</td>
</tr>
<tr>
<td>Human T-lymphotrophic virus type-1 (HTLV-1)</td>
<td>iPCR</td>
<td>d</td>
</tr>
<tr>
<td>Human papillomavirus type-16 (HPV-16)</td>
<td>Alu-PCR</td>
<td>e</td>
</tr>
</tbody>
</table>

**Transposable elements**

| Ds element | TAIL-PCR | g |
| DTph1 | iPCR | h |
| IS30 | iPCR | i |
| P element | iPCR | j |
| T-DNA | TAIL-PCR | k |
| Tc1-like | Splinkerette-PCR | l |
| Ta5 | iPCR | m |
| Ta55 | iPCR | n |
| T01 | TAIL-PCR | o |

TABLE 30.2

<table>
<thead>
<tr>
<th>Principle Based on</th>
<th>Methods</th>
<th>DNA Digestion</th>
<th>DNA Ligation</th>
<th>Amount of DNA Used</th>
<th>Primers on</th>
<th>Step to Optimize</th>
<th>Source of Nonspecific Products</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inverse</td>
<td>iPCR (Figure 30.1)</td>
<td>+</td>
<td>+ (self-ligation)</td>
<td>D: 0.2–10 µg L: 0.2–5 ng/µl</td>
<td>Integrant</td>
<td>1. Relative high DNA sample amount 2. DNA concentration of self-ligation</td>
<td>1. Unintegrated DNA 2. Unligated DNA</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>PI-PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LR-iPCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L1-PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRS-PCR</td>
<td>Novel Alu-PCR (Figure 30.2)</td>
<td>-</td>
<td>-</td>
<td>P: 10–106 ng</td>
<td>Integrant 1. IRS orientation 2. Distance to IRS</td>
<td>Inter-IRS amplification</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LINE-PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B1-PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LA-PCR</td>
<td>Vectorette-PCR (Figure 30.3)</td>
<td>+</td>
<td>+</td>
<td>D: 0.5–2 µg L: 0.5–2 µg</td>
<td>Integrant 1. Ligation efficiency 2. Adapter unit design</td>
<td>End-repair priming</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MTV-PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LDV-PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Splinkerette-PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP-PCR</td>
<td>TAIL-PCR (Figure 30.4)</td>
<td>-</td>
<td>-</td>
<td>P: 20–150 ng</td>
<td>Integrant 1. Primer set design 2. Arbitrary primer</td>
<td>Arbitrary priming</td>
<td>High</td>
<td></td>
</tr>
</tbody>
</table>

Note: D: DNA digestion reaction; L: ligation reaction; P: PCR reaction.
sensitivity. They have already been applied successfully to identify integration sites of viruses or transposable elements (Table 30.1). In this chapter, we have reviewed four different methods for the PCR-based cloning of flanking DNA sequence. Each technique has its own challenges in areas such as specificity and throughput (Table 30.2). Among the techniques introduced above, IRS-PCR and TAIL-PCR are highly recommended. They are a relatively sensitive and straightforward protocols requiring neither genomic DNA digestion nor ligation.

ACKNOWLEDGMENTS

We express our thanks to Thomas Weissensteiner for his critical reviewing of this manuscript. The research on HBV IHJ was partially supported by the grants from the National Science Council and the Minister of Education, Republic of China. S.J.L. received research awards from the National Science Council.

REFERENCES


I. INTRODUCTION

Inverse PCR\(^1\text{-}^3\) is a useful technique for supplementing conventional PCR screens that often yield only a substantial fragment of the gene of interest. An inverse PCR strategy allows the user to “walk” both upstream and downstream of a known DNA sequence in order to obtain additional sequence and facilitates amplification of unknown DNA flanking sequences without the labor involved in constructing and screening libraries. The main steps are as follows (Figure 31.1):

1. Purify and cut the genomic DNA sample with appropriate restriction enzymes (EcoR1 and Hind III in this example). Ligate the fragments into circular templates.
2. Perform inverse PCR reactions using primer sequences within the region of known sequence.
3. After the inverse PCR product has been cloned and sequenced, a full-length contiguous fragment may be amplified by conventional PCR using primer sequences at or adjacent to the restriction sites from a genomic DNA template. Search for new restriction sites within this latest cloned stretch of DNA.
4. Cut the genomic DNA sample with the second restriction enzyme. This creates additional fragments that will serve as template for the inverse PCR reaction and facilitates amplification of novel sequences both upstream and downstream of the known sequence (Figure 31.1, second set of Hind III sites and primers denoted by gray arrows).
5. Obtain additional clones by inverse PCR. The overlap between the new and old clones allows proper orientation of the inverse PCR fragments and prevents the creation of gaps.
in a contiguous sequence. PCR primers are designed from regions of known sequence as depicted in the diagram.

6. Amplify full-length contiguous DNA fragments by conventional PCR using primer sequences at or adjacent to the outermost restriction sites. Using this strategy, between 5 and 10 kb of novel contiguous DNA sequence information can be obtained after only two rounds of inverse PCR. Further inverse PCR cycles using fragments left and right of the starting fragment will yield more sequence. Each cycle of the strategy allows the user to “walk” both upstream and downstream of a known DNA sequence.
Identification of flanking sequences on either side of a known sequence can be challenging with PCR-based methods that employ random priming or the use of adapter sequences (anchored and RACE PCR) and frequently yield a low signal-to-noise ratio. By contrast, inverse PCR uses two sequence-specific primers and, therefore, generally yields a high signal-to-noise ratio. In addition, the inverse PCR primer pair can be used to amplify multiple circularized gene fragments that contain overlapping restriction sites. However, the inverse PCR works best for DNA fragments within a limited size range (see Section IV, Discussion, at the end of this chapter).

In this chapter we describe protocols for using an inverse PCR to clone contiguous sequences upstream and downstream of a known DNA sequence. Although the protocol is described for use with *Tetrahymena* DNA templates, it is readily applicable to any other genomic DNA. Recent use of this technique includes studies of Y-chromosome elements in plants, histone genes, proviral DNA, and DNA deletion elements.

**II. MATERIALS AND METHODS**

**A. PRIMERS**

We have used a variety of primers for amplification of ciliated protozoan sequences. The primers were designed to amplify from circular templates and therefore have their 3’ ends directed away from each other. These primers were generally 21- to 24-mers and 50% GC-rich with a $T_m$ between 60 to 68°C. A stock solution at 100 µM was prepared by resuspending the primers in sterile distilled water or TE buffer (1X).

**B. DNA TEMPLATES**

*Tetrahymena* genomic DNA was prepared by using a commercially available DNA Extraction Kit (Stratagene) with a final ethanol precipitation step. DNA was digested in small tubes, each containing 30 µg genomic DNA, 30 units of EcoR1 or Hind III, enzyme buffer (supplier’s directions), and distilled water to 50 µl. The tubes were incubated at 37°C for 1 h. Restriction digest products were purified using a Chromospin column (Clontech). Alternatively, DNA digestion products were purified by gel electrophoresis, which was also useful for the identification of components of optimum size for circularization (see Section IV, Discussion, at the end of this chapter). For gel purification, the band of interest was excised with a clean, sharp razor blade and purified using a Gel Extraction Kit (Qiagen).

**C. CIRCULARIZING DNA TEMPLATES**

Serial dilutions of the digested, purified DNA were prepared. A range of ligation reactions with varying concentrations (10 ng to 1 µg) of the purified DNA digests were set up in small tubes. Each reaction mixture contained the appropriate amount of diluted DNA and 6 Weiss units of T4 ligase in ligation buffer consisting of 66 mM Tris-HCl (pH 7.6), 6.6 mM MgCl$_2$, 0.1 mM ATP, 0.1 mM spermidine, 10 mM DTT, and stabilizers. The tubes were incubated at 16°C for 60 min. The circularized fragments were purified by using a CHROMASPIN column.

**D. INVERSE PCR**

Serial dilutions of the circularized DNA described in section C were prepared, and PCR reactions were set up using varying concentrations of circularized DNA. Each PCR reaction contained:
• 5 µl 10x Advantage 2 PCR Buffer (Clontech Laboratories)*
• 1 µl 50x dNTP mix**
• 1 µl (each) primers (0.5 µM each)
• 5 ng to 0.5 µg circularized DNA
• 1 µl 50x Advantage 2 Polymerase Mix (Clontech Laboratories)***
• PCR Grade water to 50 µl

The PCR program consisted of:

• pre-PCR hold at 94°C for 10 to 15 sec
• 18 rounds of denaturation at 94°C for 12 sec
• annealing/extension at 65°C for 3 to 5 min (1 min/kb)
• 12 cycles of denaturation at 94°C for 12 sec
• annealing/extension at 65°C for 3 to 5 min (1 min/kb, adding time increments of 12 sec per cycle)
• final extension at 72°C for 10 to 15 min

E. ANALYSIS OF PCR PRODUCTS

PCR products were analyzed on 1% agarose gels in TAE buffer using standard electrophoresis protocols. To confirm that the PCR product was the result of authentic amplification from a circular template, an aliquot of the PCR product was digested with the same enzyme (EcoR1 or Hind III) that generated the original template DNA, and the digestion products were analyzed by gel electrophoresis. The amplified products were cloned and sequenced.

III. RESULTS

Circularized EcoR1- or Hind III-generated genomic DNA fragments yielded sequence information upstream and downstream of the original known DNA sequence (Figure 31.1). PCR amplification of sequences in circularized EcoR1-digested genomic DNA yielded sequences predominantly downstream of the PCR priming site (Figure 31.1). Amplification of circularized Hind III-digested fragments produced sequences upstream and downstream of the PCR priming site (Figure 31.1). Digestion of an aliquot of the amplified product with either EcoR1 or Hind III produced two fragments as expected for authentic amplification from a circular template (Figure 31.1).

IV. DISCUSSION

In this chapter we have described a general method that can be used to apply the inverse PCR technique to clone contiguous sequences upstream and downstream of a known DNA sequence. Although the technique has been described for the amplification of *Tetrahymena* sequences, it is readily adaptable for use with any genomic DNA template.

A suitable restriction enzyme should generate fragments that are 2 to 3 kb in length and contain a four-base overhang to facilitate ligation. Select an enzyme that has a recognition site predicted to frequently occur within the template DNA based on GC and AT content. For example, in *Tetrahymena*, the AT content of genomic DNA is high, and enzymes such as EcoRI and Hind III,
which cut at AT-rich sites, yield fragments in the desired size range. In contrast, enzymes such as BamH1, which cut at GC-rich sites, yield fragments that are too long for efficient use in inverse PCR. Hybridization blotting can be used to confirm the identity of the restriction fragments.

The $T_m$ of the primers should not differ by more than 3 to 5°C from each other. If the $T_m$ of the primers varies by 5 to 10°C, use a two-step PCR reaction from the highest to the lowest $T_m$ value. For example, in a two-step PCR reaction in which the $T_m$ of one primer is 68°C and the $T_m$ of the second primer is 62°C, start with a cycle consisting of a 94°C denaturation step followed by a 68°C combined annealing/extension step. Subtract 0.2°C from the annealing/extension step in each subsequent PCR cycle. A 5-sec time increment should be added to the annealing/extension step of each subsequent cycle to compensate for a slight decrease in the DNA polymerase’s rate of nucleotide incorporation as the reaction progresses. The last of 30 cycles consist of a denaturation step at 94°C, followed by combined annealing/extension step at 62°C.

Many of the problems encountered with inverse PCR can often be traced to characteristics of the template DNA. In creating the circular DNA template, there is competition between concatamer formation and circularization of DNA fragments. The optimum DNA concentration that promotes circularization varies with the length of the DNA to be circularized and must be determined empirically for each template. In general, the optimum size range for efficient circularization is between 2 to 3 kb. Fragments outside this range fail to circularize efficiently. In order to lower the high melting temperatures required for denaturation of closed circular DNA, the DNA can be linearized at a site between the 5’ ends of the primers. Alternatively, a PCR buffer containing DMSO and glycerol (5%) as cosolvents can be used to overcome difficult secondary structure restrictions.

Although inverse PCR products are derived from monomeric circular templates, the presence of concatamers, even at low DNA concentrations, can yield unwanted amplification of noncontiguous sequences. A quick and effective way of distinguishing between an inverse PCR product and a concatamer-based amplification is to digest an aliquot of the amplified product with the same restriction enzyme used to generate the genomic template fragments. An authentic inverse PCR product contains a single internal restriction site corresponding to the original template-generating enzyme. Restriction digestion using this enzyme would therefore cleave the PCR product into two subfragments that, in most cases, are readily resolvable by agarose gel electrophoresis. However, if the restriction site is equidistant from both primers, both subfragments will have the same molecular size. The sum of the molecular sizes of the two subfragments should equal the size of the original inverse PCR product. A restriction digest of a concatamer-based amplification will always yield in excess of two subfragments due to the presence of multiple internal cleavage sites. If the inverse PCR generates multiple amplification products, each band should be excised and gel purified separately. Aliquots of the gel-purified DNA can then be used for restriction digestion analysis. This method provides a simple means of verifying an inverse PCR product prior to cloning and sequencing.

ACKNOWLEDGMENTS

The original research on the use of inverse PCR with Tetrahymena DNA was supported by grants MCB 9808301 and MCB 0130624 from the National Science Foundation to RHG.

REFERENCES


I. INTRODUCTION

Genome walking is indispensable to extensive structural analysis of large gene segments. A number of genome walking methods are available to the researcher — including iPCR,\textsuperscript{1,2} cassette ligation-anchored PCR,\textsuperscript{3} vectorette PCR,\textsuperscript{4} panhandle PCR,\textsuperscript{5,6} Alu-PCR,\textsuperscript{7} thermal asymmetric interlaced PCR (TAIL-PCR),\textsuperscript{8} and EPTS/LM-PCR\textsuperscript{9} (see also the review by Eric Ka-Wai Hui and Szecheng J. Lo in Chapter 30, and Chapter 31 on iPCR by Jorge A. Garcés et al.). However, the usefulness of these techniques has been hampered by a combination of (1) limited sensitivity/speciﬁcity, (2) a short walk range, (3) a requirement for restriction enzymes and advantageous restriction maps, (4) a requirement for efﬁcient ligation, (5) the necessity of molecular cloning, and (6) low throughputs or low rates of success.

Universal Fast Walking (UFW)\textsuperscript{10} was earlier proposed to solve these limitations since the method is characterized by high speciﬁcity and sensitivity, and by general operational efﬁciency related to an independence from restriction enzymes, ligases, and molecular cloning. These reactions also routinely generated DNA ﬁngerprints as a quality control.

Following from that method, I have developed Long Universal Fast Walking (L-UFW) for achieving signiﬁcantly longer genome walks without compromising the compactness, procedural simplicity, or reliability of the earlier system. There remain no more than four direct reagent additions, with a total reaction volume not exceeding 50 µl. In this example, I describe the use of L-UFW for closing a sizeable sequence gap found in the \textit{Drosophila} genome project.

II. MATERIALS AND METHODS

A. DNAs and Primers

Genomic DNAs were extracted from \textit{Drosophila} adults by a previously described LiCl technique.\textsuperscript{11} Custom oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, IA) or
Sigma-Genosys (The Woodlands, TX), and the 1-kb DNA marker ladder was obtained from Invitrogen (Carlsbad, CA).

**B. The L-UFW Strategy**

Long Universal Fast Walking (see Figure 32.1 for the mechanism, Table 32.1 for the protocol) is a one-tube process completed with one temperature cycle program. After synthesis of a first complementary strand with primer 1, any nonincorporated primer is destroyed with exonuclease I (exoI) (United States Biochemical, Cleveland, OH). After denaturation, strands are annealed with primer 2, containing a random segment at its 3' end. The first strand is then exoI-digested back to its branch point with the distally annealed primer 2. The polymerase catalyzes a repair that converts the first-strand end-sequence using the 5' (nonrandom) portion of primer 2 as a template. Denaturation of strands allows a first-strand lariat to form by intrastrand annealing between the tag complement, segment 2', and a copy of the tag, segment 2, near the walk origin. Repair of the lariat stem is followed by PCR amplification, using the Takara long-and-accurate polymerase blend (PanVera, Madison, WI). Reactions were run in the T-gradient Cycler (Biometra, Göttingen, Germany), a device permitting releasable interim pauses and finely adjusted ramping.

For each reaction, reagent premixes were added to the starting mix in 5 μl increments at each step, until reaching the final reaction volume of 50 μl. For increased throughput, reaction sets were run in 96-well polypropylene trays (Corning Inc., Corning, NY) throughout.

Both exoI and the polymerase were active in the same PCR buffer supplied by the manufacturer; the presence of one enzyme was not observed to interfere with the activity of the other. Nesting of primers 3 and 4 (Figure 32.1) and the use of cold-starting — i.e., making additions below the

---

**FIGURE 32.1** Diagram of L-UFW, a one-tube process with a longer range than the original UFW. Unknown flanking DNA is shown as a wavy line. Sequences of L-UFW primers, numbered in temporal order, are derived precisely from their respective target segments, where primer 2 also contains a random terminal sequence. In some experiments this random motif is split by a deoxy-, abasic dinucleotide (upper left of the figure) to reduce the formation of non-first-strand lariat precursors. The starting orientation of the interval from segment 2 to segment 3 is ultimately reversed by these reactions.
active temperature of the polymerase, then reinserting the reaction tray when the cycler has arrived at high temperature — were employed to maintain specificity.

C. Gel Analysis and Sequencing

L-UFW amplicons were visualized by gel electrophoresis in 1% agarose and ethidium bromide staining. For sequencing, products were prepared with QIAquick (Qiagen Inc., Valencia, CA) or Montage (Millipore, Bedford, MA) PCR filtration units. DNA sequence was determined by nucleotide chain termination techniques.\[^{12}\]

**TABLE 32.1**

**L-UFW: Reaction Conditions and Cycler Program**

<table>
<thead>
<tr>
<th>First-Strand Synthesis</th>
<th>First-Primer Destruct</th>
<th>Second-Primer Random Binding</th>
<th>Branched-End Conversion</th>
<th>Long-PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 µl genomic DNA (50–100 ng)</td>
<td>Add a pre-mix of: 0.5 µl exoI (5 U)</td>
<td>Add a pre-mix of: 1.0 µl (15 pmol) primer 2</td>
<td>Add a pre-mix of: 1.0 µl exol (10 units) primer 3</td>
<td>Cold-start</td>
</tr>
<tr>
<td>3.0 µl 10X PCR buffer</td>
<td>0.5 µl 10X PCR buffer</td>
<td>0.5 µl 10X PCR buffer</td>
<td>0.5 µl 10X PCR buffer</td>
<td>1.0 µl (10 pmol) primer 4</td>
</tr>
<tr>
<td>4.8 µl 10 mM dNTP mix</td>
<td>4.0 µl dH₂O</td>
<td>3.2 µl 10 mM dNTP mix</td>
<td>3.5 µl dH₂O</td>
<td>0.5 µl 10X PCR buffer</td>
</tr>
<tr>
<td>0.6 µl (6 pmol) primer 1</td>
<td></td>
<td>0.3 µl dH₂O</td>
<td></td>
<td>2.5 µl dH₂O</td>
</tr>
<tr>
<td>0.5 µl Takara LA-Taq (2.5 units)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.1 µl dH₂O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cold-start 95°C, 3 min 53°C, 30 sec 68°C, 4 min Pause at 37°C</td>
<td>37°C, 30 min</td>
<td>Cold-start 94°C, 5 min 68°C, 1 sec 20°C, 1 sec, ramping 0.02°C/sec</td>
<td>37°C, 45 min 68°C, 15 min 80°C, 15 min 94°C, 3 min 65°C, 1 sec 45°C, 1 sec, ramping 0.03°C/sec</td>
<td>Cold-start 95°C, 2.5 min, then run 36 cycles of: 98°C, 20 sec 49°C, 30 sec 68°C, 10 min, adding 11 sec/cycle</td>
</tr>
<tr>
<td></td>
<td>Pause at 4°C</td>
<td>68°C, 2 min Pause at 4°C</td>
<td>68°C, 10 min</td>
<td>Final extension for 68°C, 10 min</td>
</tr>
<tr>
<td>Comments: First-strand length is governed by the extension time. Annealing temperature can be varied for the particular primer. Long-distance polymerase is essential. Cycler steps for the entire protocol are contained within a single cycler program.</td>
<td>Comments: Digest first primer. For high throughput, reagent additions may be from pre-mixes.</td>
<td>Comments: Random annealing of 5’-tagged primer. Prior heat-inactivation of exol is unnecessary due to cold-start. Ramping shown is downward from 68–20°C, since ramp rates appear on the program line for the destination temperature.</td>
<td>Comments: First-strand trimming and sequence conversion at the 3’ end of the first strand. Intramolecular annealing of the first strand, with extension, to form a lariat.</td>
<td>Comments: Amplification and first-strand extension times are mutually adjustable for longer walks. Nesting eliminates lariat stem. Annealing temperature is adjustable for Tm.s. Reported settings are for a <em>Drosophila</em> genome walk.</td>
</tr>
</tbody>
</table>
III. RESULTS AND DISCUSSION

A. DIRECT GENOME WALKING AND DNA FINGERPRINTING

Much attention was paid to operational and spatial economy at all points. The restriction digests and DNA ligations that characterize other walking methods were thus eliminated, as well as the dilution/reconcentration cycles that often accompany those steps. Of further interest was the reduction of the material losses that hinder many of the other methods. L-UFW is based strictly on component additions, in which none of the starting material is removed during the procedure. Consequently, this walking technique is in principle within the same range of sensitivity as direct long-PCR. Note also that the L-UFW primer nesting, in addition to enhancing specificity, favors the removal of the lariat stem during amplification (Figure 32.1).

Requiring only nanogram amounts of *Drosophila* genomic DNA, L-UFW agarose gel analysis (Figure 32.2) from transposon mapping studies showed consistent production of sequence-grade (Figure 32.3) fragments in a size range beyond that of standard PCR. These results indicate that L-UFW might be suitable for the analysis of very large genome intervals, conceivably with a walk-range approaching the direct PCR limit for which the associated long-distance polymerase is qualified. Accordingly, L-UFW was applied to a multikilobase gap present in the *Drosophila* genome project (Figure 32.4). Again, starting with only nanogram quantities of template, sufficient product was obtained for direct sequencing of the amplicon to fill the gap (Figure 32.4A); sequence was obtained without further isolations such as gel extractions. The agarose gel patterns (Figure 32.4B), while rather complex, were reproducible and useful as DNA fingerprints for quality control. This banding complexity is probably due to (1) partial site preferences in the distal binding of primer 2 and (2) pause sites encountered during the first 37°C step of L-UFW, a temperature at which the thermophilic polymerase would be only partly active. Fingerprint complexity was not adverse to sequence determination, consistent with the notion that all of the reaction products in a given fingerprint should derive from the same walk origin, and therefore share its flanking sequence.

B. POTENTIAL APPLICATIONS FOR L-UFW

In applications such as insertional element (e.g., transposon) mapping, currently the UFW and L-UFW techniques appear to be methods of choice, owing to their reliability, along with their compactness, minimal handling, and quality controls. Since the direct-additions regime is among

![Figure 32.2](www.taq.ir) DNA fingerprints of *Drosophila* genome walks from P-transposon walk origins. Gel patterns are particular to each flanking region. At high throughput, a UFW/L-UFW reaction set is run from beginning to end without changing the 96-well PCR tray. Lanes M: 1-kb ladder makers.
the simplest of reagent handling formats, the methods reported here should be highly amenable to miniaturization and automation, making the UFW family of techniques suitable for large-scale mutational analysis.

Significantly, L-UFW typically circumvents not only restriction-enzyme-based cloning but all molecular cloning. This technique is therefore likely to be effective for determining genomic sequence for a range of intervals that are not sequenceable due to their recalcitrance to the standard library cloning strategies. L-UFW should also be desirable for such uses as evaluating the integration sites of gene therapeutic vectors and of biomedically important agents such as retroviruses.

A potential difficulty that L-UFW might encounter is with a walk origin representing an exact copy amongst multicopy (though nontandem) genes, since in that case direct sequencing through the mixed flanking DNA would be difficult or impossible to interpret. Possible solutions would be to combine L-UFW with molecular cloning or gel-band extraction (recalling that L-UFW yields reproducible DNA fingerprints) to generate clean sequence. Of course, cellular gene copies that are at all dissimilar provide the opportunity to engineer L-UFW primers just sufficient for selective amplification.

In conclusion, this study demonstrates the ability of Long Universal Fast Walking to resolve appreciable sequence gaps, independent of the local restriction map, and without relying on DNA ligation or molecular cloning. Like UFW, L-UFW yields a reproducible DNA fingerprint as a useful control that is specific to the genomic or transposon flanking region. In addition to being highly effective for rapid transposon mapping, the UFW and L-UFW systems provide a general strategy for obtaining genomic sequence from regions that may be rare or nonexistent within the conventional whole genome libraries. By design, the walk range of this strategy is closely related to the limits of the amplifying polymerase and should therefore increase along with the continued advances in the enzyme technology.
ACKNOWLEDGMENTS

This study was funded by the NIGMS of the National Institutes of Health (USA), grant #R37 GM28669-24, and I am grateful to laboratory chief William M. Gelbart for his support of this work.
REFERENCES

33 Cloning of Unknown Virus Sequences by DNase Treatment and Sequence-Independent Single Primer Amplification

Tobias Allander, Suzanne U. Emerson, Robert H. Purcell, and Jens Bukh

CONTENTS

I. Introduction ................................................................................................................. 337
II. Example ..................................................................................................................... 339
   A. Key Reagents ......................................................................................................... 339
      1. Adapters ........................................................................................................... 339
      2. Enzymes .......................................................................................................... 340
   B. Experimental Protocol .......................................................................................... 340
      1. Sample Pretreatment ....................................................................................... 340
      2. DNA Protocol: Extraction and Synthesis of a Second Strand of DNA .......... 340
      3. RNA Protocol: Extraction and cDNA Synthesis ............................................ 341
      4. Restriction Digestion and Amplification ......................................................... 341
      5. Analysis of PCR Products .............................................................................. 342
      6. Sequence Analysis .......................................................................................... 343
   C. Interpretation of Data ............................................................................................ 343
      1. Contamination .................................................................................................. 343
      2. Nonviral sequences ......................................................................................... 343
      3. Viral Sequences .............................................................................................. 343
      4. Sequences with No Homology to Database Sequences ................................... 343
      5. Disease Association ....................................................................................... 344
III. Discussion ................................................................................................................. 344
References ...................................................................................................................... 344

I. INTRODUCTION

For many human diseases infection by unrecognized viruses is a possible etiology. Moreover, the identification of new viral species is important for understanding our normal physiology and microbiological environment. Traditionally, tissue culture has been the basic tool for isolation and characterization of viruses. However, many viruses are difficult or impossible to grow in cell culture.
A major breakthrough for the molecular approach to virus discovery was the cloning of hepatitis C virus from chimpanzee plasma in 1989. The cloning of viral nucleic acids directly from plasma was particularly encouraging since frozen serum and plasma samples derived from patients with medical disorders of unknown etiology are widely available.

The main obstacle to the identification of viruses in serum and plasma by molecular techniques is the extremely small quantity of viral nucleic acid present, in combination with a relatively high level of host genomic DNA. Therefore, molecular virus discovery methods generally involve amplification as well as selection of the target sequences (reviewed in Muerhoff et al. and Kellam). Three basic approaches for the identification of completely unknown viral sequences have been applied:

1. Library construction and colony screening (immunoscreening or differential hybridization), with or without prior nonspecific nucleic acid amplification (e.g., sequence-independent single-primer amplification, SISPA)
2. Subtractive hybridization (e.g., representational difference analysis — RDA)
3. DNase treatment followed by nucleic acid extraction and amplification (e.g., DNase-SISPA)

All three methods have been applied to the discovery of new virus species. The first approach has yielded medically important discoveries such as the cloning of hepatitis C and E viruses. However, the screening procedure requires carefully selected pairs or sets of samples and is very time consuming. The second approach (RDA) showed very impressive performance for virus discovery in the original publication and was therefore received with great enthusiasm by virologists. It also requires matched sample pairs. It is less time consuming than library screening, but still relatively labor-intensive. Although some virus species have been discovered by the use of RDA, the initial great expectations for the method have not been met. A likely explanation for this is that optimal subtractive hybridization requires extremely well-matched sample pairs. These are difficult to obtain since serum or plasma samples collected from the same individual are actually often quite heterogeneous in nucleic acid content.

In order to find a less labor-intensive and more reproducible method for virus discovery, we recently developed the DNase-SISPA method. It can be applied to cell-free liquids such as serum, plasma, or cell culture supernatants. The basic principle of this method is that the amount of nonviral nucleic acids in the sample is reduced significantly by filtration and DNase treatment (Figure 33.1). Viral nucleic acids are protected within the virus particle at this stage and are then extracted, converted to double-stranded DNA, and digested with a restriction enzyme that cuts relatively frequently (typically an enzyme that recognizes a four-base pair sequence and cuts asymmetrically). A matching adapter is ligated to the ends of the restriction fragments and the fragments are amplified by PCR with an adapter-specific primer. The PCR products are visualized on a gel and any virus present at sufficient titer will produce distinct DNA bands that can be cloned and sequenced, while a virus-free sample usually results in a faint DNA smear.

The major advantage of this method is the limited effort required for investigating each sample. The protocol readily allows analysis of 10 samples in 3 d by a single scientist, and the method may be adapted further for more efficient screening. Importantly, matched pre- and postinfection samples are not required. Thus, for the first time, screening of multiple samples for unknown viruses is feasible. Another major advantage is that the method gives reproducible results. DNA and RNA viruses having a titer of approximately 10^6 genome copies/ml or greater in serum samples were repeatedly detected in titration experiments with different known viruses. Finally, the method facilitates the sequencing of a large portion of the viral genome directly upon discovery if the virus titer is high enough. While evaluating the method, we detected two unknown bovine parvovirus species in the commercial calf serum used for dilution of the samples. This confirmed that DNase-SISPA is a useful method for virus discovery and also illustrated that this method can be used to search for viruses without any detailed knowledge of the sample.
Cloning of Unknown Virus Sequences

Serum/plasma/cell culture supernatant
  ↓
  Filtration
  ↓
  DNase treatment
  ↓
  DNA and RNA extraction
  ↓
  dsDNA synthesis
  ↓
  Restriction enzyme digestion
  ↓
  Ligation of adaptor to resulting DNA fragments
  ↓
  PCR amplification with adaptor-specific primer
  ↓
  Gel analysis of PCR products
  ↓
  Cloning and sequencing of defined bands
  ↓
  Database homology search

FIGURE 33.1 Principle of the DNase-SISPA method. See text for details.

The major limitation of DNase-SISPA is the sensitivity. Viruses with a titer lower than $10^6$ genome copies/ml are unlikely to be detected, and blood-borne viruses frequently have such low titers. On the other hand, viral titers vary greatly from sample to sample, and the limited sensitivity can therefore be partly overcome by analyzing more samples. In particular, samples collected during acute disease may have high virus titers. Finally, it is not known how the sensitivity of DNase-SISPA compares with that of the other available discovery methods. The reproducible detection limit for virus discovery in serum or plasma by library screening or RDA has not been reported.

II. EXAMPLE

A. Key Reagents

1. Adapters

The adapter NCsp was made by annealing two 5'-nonphosphorylated (the normal format for synthetic oligonucleotides) and HPLC-purified oligonucleotides (NBam24, AGG CAA CTG TGC TAT CCG AGG GAG and NCsp11, TAC TCC CTC GG). The oligos were reconstituted to 250 µM in TE buffer. The annealing reaction was set up as follows:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NBam24 (250 µM)</td>
<td>20 µl</td>
</tr>
<tr>
<td>NCsp11 (250 µM)</td>
<td>20 µl</td>
</tr>
<tr>
<td>NaCl (5M)</td>
<td>1 µl</td>
</tr>
<tr>
<td>TE</td>
<td>9 µl</td>
</tr>
<tr>
<td>Total volume</td>
<td>50 µl</td>
</tr>
</tbody>
</table>

The annealing reaction was incubated in a thermocycler (Perkin–Elmer Thermal Cycler 480) at 65°C for 10 min followed by a gradual decrease from 65°C to 10°C over 120 min (thermocycle file). The resulting 100-µM stock solution and aliquots of diluted 10-µM working solutions were stored at −20°C.

www.taq.ir
2. Enzymes

Most, if not all, enzymes for molecular biology contain DNA derived from the microorganisms in which they are produced, and possibly from other sources. Since all DNA introduced into the sample between the DNase treatment and the restriction digest step is amplified, it is crucial to minimize the input of DNA into the sample. Too much exogenous DNA in the sample can result in a broad smear of DNA in the gel following PCR, which is likely to reduce the sensitivity for detecting viruses. More frequent, however, were distinct bands in the gel following PCR (false positive results). These DNA products could usually be identified by sequencing as representing bacterial DNA and their origin could be traced to the reagents. We did not experience either of these problems with the protocol for detecting DNA viruses. For the RNA virus protocol, however, reagent DNA (and possibly RNA) was a problem. This is most likely because a number of enzymes are needed for the cDNA synthesis steps unique to the RNA virus protocol, and contamination of only one of them is enough to cause problems. While low DNA-content Taq polymerase is commercially available, it is of little help to solve this problem, and low DNA-content cDNA synthesis reagents are currently not available. The only solution we found was to test a number of reagents (and batches) by performing the complete DNase-SISPA procedure on a known virus-positive serum and on a virus-negative serum to simultaneously test sensitivity and background. The negative sample should generate a faint smear but no bands after electrophoresis of PCR products on gels. Unfortunately, the supplier discontinued the reagents that gave the best results (cDNA synthesis module, Amersham). The protocol for cDNA synthesis below is therefore to be seen only as an example.

B. Experimental Protocol

1. Sample Pretreatment

For each serum or plasma sample, two 50-µl aliquots were analyzed (for DNA and RNA virus detection, respectively). Each aliquot was diluted with 150 µl H₂O and filtered through a 0.22-µm spin filter (Ultrafree MC, Millipore) at 2000 rpm in a microcentrifuge for 5 min. In case some of the sample did not pass through the filter, it was transferred to a new filter and centrifuged for an additional 5 min. For some samples, several filter changes were necessary. After filtration, 100 U (10 µl) of DNase I (Stratagene) was added and samples were incubated for 2 h at 37°C. Thereafter, DNA was extracted from one aliquot and RNA was extracted from the other, and subsequently converted into double-stranded DNA. Thus, one aliquot was processed according to point 2 below, and the other according to point 3, before analysis of both aliquots continued under points 4 to 6.

2. DNA Protocol: Extraction and Synthesis of a Second Strand of DNA

DNA was extracted with a QIAamp blood mini kit (QIAGen) following the manufacturer’s instructions. The DNA was eluted in 50 µl of elution buffer. A second strand of DNA was synthesized by adding the following reagents directly to the DNA eluate:

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecopol buffer (New England Biolabs)</td>
<td>6.0 µl</td>
</tr>
<tr>
<td>dNTP (10 mM each)</td>
<td>0.5 µl</td>
</tr>
<tr>
<td>Random hexamers (10 µM)</td>
<td>1.0 µl</td>
</tr>
<tr>
<td>H₂O</td>
<td>6.5 µl</td>
</tr>
<tr>
<td>3'-5' exon–Klenow polymerase (5 U/µl)(New England Biolabs)</td>
<td>1.0 µl</td>
</tr>
<tr>
<td>Total volume (approximate)</td>
<td>60.0 µl</td>
</tr>
</tbody>
</table>

The reaction mixture was incubated for 1 h at 37°C followed by heat inactivation at 72°C for 10 min.
3. RNA Protocol: Extraction and cDNA Synthesis

RNA was extracted with the Trizol reagent (Invitrogen). The sample was mixed with 1 ml Trizol and extraction proceeded according to the manufacturer’s instructions. At the isopropanol precipitation step, 1 µl of 20 mg/ml glycogen (Roche) was added as a carrier. Extracted RNA was dissolved in 12 µl of 10 mM dithiothreitol (DTT, Promega) and 0.4 U/µl recombinant RNasin (Promega) in RNase-free water.

All extracted RNA was converted to double-stranded cDNA by using a cDNA synthesis kit that was first evaluated for contaminating DNA (see text above). A lower concentration of enzymes than recommended by the manufacturers was generally used. The following is an example based on a cDNA synthesis kit from Promega:

<table>
<thead>
<tr>
<th>Component</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA</td>
<td>12.0 µl</td>
</tr>
<tr>
<td>Random hexamers (10 µM)</td>
<td>1.0 µl</td>
</tr>
<tr>
<td>5 x first strand buffer</td>
<td>4.0 µl</td>
</tr>
<tr>
<td>rRNasin (40 U/µl)</td>
<td>0.5 µl</td>
</tr>
<tr>
<td>Sodium pyrophosphate (40 mM)</td>
<td>2.0 µl</td>
</tr>
<tr>
<td>Avian myeloblastosis virus (AMV) reverse transcriptase (10 U/µl)</td>
<td>0.6 µl</td>
</tr>
<tr>
<td>Total volume</td>
<td>20.1 µl</td>
</tr>
</tbody>
</table>

The reaction mixture was incubated at 37°C for 15 min immediately followed by incubation at 42°C for 45 min. The completed first-strand reaction mixture was placed on ice and the second-strand synthesis reagents were added:

<table>
<thead>
<tr>
<th>Component</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5x second-strand buffer</td>
<td>40.0 µl</td>
</tr>
<tr>
<td>H₂O</td>
<td>33.9 µl</td>
</tr>
<tr>
<td>BSA (1mg/ml)</td>
<td>5.0 µl</td>
</tr>
<tr>
<td>DNA polymerase I (5 U/µl)</td>
<td>0.8 µl</td>
</tr>
<tr>
<td>RNase H (0.5 U/µl)</td>
<td>0.3 µl</td>
</tr>
<tr>
<td>Total volume including first-strand reaction</td>
<td>100.0 µl</td>
</tr>
</tbody>
</table>

The reaction mixture was incubated in a thermocycler (Perkin–Elmer Thermal Cycler 480) at 12°C for 1 h and at 22°C for 1 h, followed by enzyme inactivation at 75°C for 10 min.

4. Restriction Digestion and Amplification

From this point, the RNA and DNA protocols are identical and the tubes can be handled in parallel. Csp6.I (Fermentas) measuring 10 units (1 µl) were added directly to each tube and the tubes were incubated at 37°C for 30 min. The DNA fragments were next extracted by the QIAquick PCR purification kit (Qiagen), following the manufacturer’s instructions except that the DNA was eluted in 25 µl of a 1:1 mix of the kit elution buffer and water. Of the extracted DNA, 7 µl was ligated to the adapter NCsp in a reaction containing:

<table>
<thead>
<tr>
<th>Component</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extracted DNA</td>
<td>7.0 µl</td>
</tr>
<tr>
<td>10 x ligation buffer (New England Biolabs)</td>
<td>1.0 µl</td>
</tr>
<tr>
<td>Adapter NCsp (10 µM)</td>
<td>2.0 µl</td>
</tr>
<tr>
<td>T4 DNA ligase (2000 NEB units/µl or 30 Weiss units/µl)</td>
<td>0.2 µl</td>
</tr>
<tr>
<td>Total volume</td>
<td>10.2 µl</td>
</tr>
</tbody>
</table>

The reaction mixture was incubated in a thermocycler (Perkin–Elmer Thermal Cycler 480) at 4°C for 1 h, 16°C for 4 h, and 4°C overnight.
Of the ligation reaction, 2 µl was used as template for PCR, using the oligonucleotide NBam24 as a primer. The reaction was set up as follows:

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂O</td>
<td>30 µl</td>
</tr>
<tr>
<td>10 X PCR buffer II (Applied Biosystems)</td>
<td>5 µl</td>
</tr>
<tr>
<td>MgCl₂ (25 mM)</td>
<td>5 µl</td>
</tr>
<tr>
<td>dNTP (10 mM)</td>
<td>1 µl</td>
</tr>
<tr>
<td>NBam24 (10 µM)</td>
<td>2 µl</td>
</tr>
<tr>
<td>Template</td>
<td>2 µl</td>
</tr>
<tr>
<td>Total volume</td>
<td>45 µl</td>
</tr>
</tbody>
</table>

AmpliTaq DNA polymerase (Applied Biosystems) was diluted 1:10 in 0.7 x PCR buffer II. A manual hotstart procedure was applied: All tubes, including that containing the diluted enzyme, were heated to 72°C for 3 min. Diluted enzyme measuring 5 µl was then transferred to each PCR reaction mixture. After an additional 5 min at 72°C to fill in the adapter DNA and generate the primer binding sites, 40 cycles of amplification (94°C for 1 min and 72°C for 3 min) were performed in the Perkin–Elmer Thermal Cycler 480.

5. Analysis of PCR Products

PCR products were separated on a 1.5% agarose gel. Figure 33.2 shows the result of titrations of three viruses [hepatitis B virus (dsDNA), bacteriophage M13 (ssDNA), and GB virus B (ssRNA)]. If bands were not visible, the analysis was finished at this point. Visible bands were cloned and sequenced. Intense bands were cut out directly, extracted from the gel, and cloned. In order to analyze faint bands, a piece of the gel was punched out with a pipette tip and directly transferred to a new PCR reaction tube in which the fragment was reamplified in a second PCR using the primer NBam24. This procedure facilitated cloning of barely visible bands.

The adapters were removed before the fragments were cloned. If not, the repeated adapter sequence at both ends of the fragment caused DNA structure-related sequencing artifacts. Therefore, 1 µl RsaI (10 U) was added to 15 µl of the PCR product and incubated at 37°C for 1 h. This

![Figure 33.2](https://www.taq.ir) Detection levels of different virus genomes by DNase-SISPA. Hepatitis B virus (double-stranded DNA virus), bacteriophage M13 (single-stranded DNA virus), and GB virus B (single-stranded RNA virus) were diluted in 10-fold increments in bovine serum and subjected to the DNase-SISPA procedure. Products were separated on an agarose gel. The virus titer (genome equivalents/ml) is shown above each lane. M: molecular weight marker PhiX 174, HaeIII fragments. (Adapted from Allander, T. et al., A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species, Proc. Natl. Acad. Sci. U.S.A., 98, 11609, 2001.)
resulted in blunt-ended adapter-free products. The products were separated on an agarose gel, and bands of interest were cut out from the gel and extracted using standard techniques. The fragments were cloned into the vector pCR-Blunt (Zero Blunt PCR cloning kit, Invitrogen), and sequenced with M13 forward and reverse primers.

6. Sequence Analysis

Usually two clones were sequenced for each DNA band. The sequences were subjected to homology search with BLAST (via http://www.ncbi.nlm.nih.gov/blast/). Each sequence was analyzed for nucleotide sequence homology (BLASTN) by searching the standard nonredundant databases as well as the high-throughput genomic sequence database. Deduced amino acid sequence homology was analyzed with the translated BLAST search (BLASTX).

C. INTERPRETATION OF DATA

1. Contamination

Standard precautions to avoid PCR product carryover are essential. Most procedures are performed prior to PCR, and all relevant reagents and equipment must be kept separate from experiments performed after PCR. It is also recommended that reasonable precautions be taken to avoid contamination of the samples by any exogenous DNA, e.g., bacteria in the environment. Aerosol-resistant pipette tips and sterile plasticware are recommended when applicable.

2. Nonviral sequences

We have analyzed the composition of the PCR product smear in negative samples by sequencing random clones. Some host sequences from the serum were seen, but a large number of clones did not match in the database search when nucleotide sequences were investigated. However, homology to bacterial sequences was seen when the clones were analyzed as deduced amino acid sequences. These clones were therefore considered to be of bacterial origin, and by PCR with specific primers they could be traced to the reagents.

3. Viral Sequences

Viral sequences can be expected to be either sequences with homology to known viruses or sequences without homology to known organisms. The first group can be identified by database search. However, the two parvoviruses we identified had very low similarity to other parvoviruses at the nucleotide level. The cloned fragments of 100 to 400 nucleotides could not be identified at all by nucleotide similarity search, but some of the fragments had deduced amino acid sequences with significant similarity to parvoviruses. Thus, the database search will have to be thorough and exhaustive. For a more definitive classification of unique viral sequences, the near complete genome sequence should be determined if possible.

4. Sequences with No Homology to Database Sequences

Clones without homology to database sequences (nucleotide and deduced amino acid sequence) pose a particular problem. They are potentially interesting findings, but it may be too time-consuming to investigate all of them. For further analysis, we typically selected clones with long open reading frames, made a pair of specific primers, and analyzed for the presence of the sequence in all SISPA products of one experiment (4 to 10 samples). It can be presumed that sequences present in most or all products are unlikely to be sample-derived but are likely to be reagent-derived. For candidates that pass this test the next step should be to confirm their presence in the original sample and exclude their presence in the host genome by PCR with specific primers. The following
step could be to extend the sequence by various genome-walking techniques or by cloning additional fragments in order to classify better the identified sequence.

5. Disease Association

Great care must be taken before claiming an association between a newly discovered virus and a disease. For example, hepatitis G virus (also called GBV-C) and the TT virus were originally believed to cause hepatitis in humans since they were cloned from individuals with hepatitis. However, further studies could not confirm an association of these viruses with hepatitis. Many viruses such as the herpesviruses are widespread in the population but may still cause specific diseases in selected individuals. Finding a plausible association between a virus and a disease may therefore require thorough investigations by several research groups and may not be achievable by the researchers discovering a virus. For review and guidelines on virus-disease association, see Fredricks and Relman.

III. DISCUSSION

DNase-SISPA is a relatively simple and inexpensive method for virus discovery, allowing simultaneous analysis of multiple samples. However, establishing the method in the laboratory requires some basic investments in finding appropriate reagents, low in contaminating DNA, for the cDNA synthesis. It is necessary to establish positive controls of serum samples having relatively high titers of known viruses or cultured viruses diluted in serum. Once the method is established, approximately 10 samples can be analyzed in 3 d with the current protocol, sequencing of positive samples excluded. False positive results may occur, and it is recommended that there be good access to DNA sequencing.

DNase-SISPA can be combined with other methods for virus discovery in order to improve sensitivity. DNase-SISPA products can be cloned into a library and subjected to immunoscreening. It is also possible to combine DNase-treatment of serum with RDA. DNase-SISPA is a suitable method for rapid sequencing of moderate-sized viral genomes that have sufficient titers. If the sequence is partly or largely unknown, DNase-SISPA will facilitate sequencing by producing fragments convenient for sequencing without any time-consuming genome-walking procedures. By combining two or more restriction enzymes, overlapping fragments will be produced that facilitate sequence assembly. Finally, DNase treatment of serum is a generally useful tool for studies of viruses in serum whenever background from host DNA is a problem, for example when using degenerate PCR or anchored PCR. In addition to virus discovery and sequencing, we have applied DNase-SISPA to the analysis of cell culture supernatants in order to identify viruses causing cytopathic effect in cell culture (unpublished data). This application may find future use for general screening of cell cultures for contaminating or adventitious viruses or perhaps in clinical diagnostics.

REFERENCES

Cloning of Unknown Virus Sequences

Amplification of Minute Amounts of Nucleic Acids
34 Whole Genome Amplification from Single Cells and Minute DNA Samples

Dagan Wells and Mercedes G. Bermudez

I. INTRODUCTION

Minute DNA samples are a feature of several areas of genetic testing including forensics, the study of ancient DNA, analysis of microdissected tissue, and certain forms of prenatal diagnosis. The use of small amounts of DNA limits the number of polymerase chain reaction (PCR)-based analyses that can be carried out. In the most extreme cases, such as preimplantation genetic diagnosis (PGD), where just one cell is available for testing, only a single polymerase chain reaction (PCR) amplification can be performed per sample. Genetic investigations may also be limited in cases where the DNA sample is unique and cannot be replenished, for example, when samples were obtained from deceased individuals or from small biopsies taken during surgical procedures.

The most straightforward means of maximizing the genetic information obtained from a small sample is to simultaneously amplify and analyze several DNA fragments in each PCR. This process, known as multiplex PCR, achieves concurrent amplification of several distinct genetic loci by adding extra pairs of primers to the PCR mixture. However, some sets of primers are incompatible and cannot be combined in the same reaction. Incompatibility can occur if the optimal reaction conditions for two sets of primers are too dissimilar. Alternatively, the primers for different loci may anneal to each other or interact with the products that they produce in such a way that
amplification is inhibited. Such problems are increasingly likely to be encountered as more sets of primers are added to an amplification reaction. These factors, along with competition between amplified loci for PCR reagents, make it difficult to efficiently amplify more than a handful of distinct loci in a single reaction.

As with all PCR methods, the original sample (e.g., the single cell or aliquot of DNA used to set up the PCR) is lost after multiplex amplification. In the case of single-cell PCR, this fact means that it is not possible to return to the sample at a later date and perform confirmatory diagnoses or other tests. Whole genome amplification (WGA) overcomes this limitation by providing a source of DNA that can be used to set up multiple independent PCRs. Unlike most PCR strategies, the amplification is nonspecific and aims to generate an increase in the copy number of all sequences in the genome, rather than a single defined fragment. Several WGA methods utilize heterogeneous mixtures of random or semirandom primers. Under appropriate reaction conditions these oligonucleotides will anneal throughout the genome and thus initiate DNA synthesis from multiple sites simultaneously.

The method known as primer extension preamplification (PEP) employs a mixture of degenerate oligonucleotides, usually 15 bp in length. This method of WGA has been applied to the analysis of multiple DNA fragments in individual cells, allowing diagnosis of genetic disorders such as Tay-Sachs, cystic fibrosis, hemophilia A, and Duchenne muscular dystrophy in a research context. PEP has also been used for the detection of genetic disorders in single cells biopsied from human embryos, thus allowing clinical PGD. However, the amplification afforded by PEP is relatively modest, and although adequate for PCR-based tests, it is insufficient for downstream applications that require larger amounts of DNA.

One of the most widely applied forms of WGA is degenerate oligonucleotide primed PCR (DOP-PCR). This method utilizes oligonucleotide primers with specified 5’ and 3’ ends separated by a short stretch of random bases. A number of PCR cycles conducted at low annealing temperatures allow the semidegenerate primers to anneal at numerous sites in the genome. Later, a change in reaction conditions allows the products of earlier cycles to be amplified exponentially (Figure 34.1). Consequently, DOP-PCR generates a much larger quantity of DNA than PEP. It has been shown that DOP-PCR performed on a single cell can provide enough DNA for more than 90 subsequent PCR amplifications. DOP-PCR has most often been used for the generation of DNA probes, usually chromosome paints made from microdissected or flow sorted chromosomes. However, there is potential for this technique to be used in other molecular genetic applications.

A method related to DOP-PCR which employs a mixture of primers that contain a specified 5’ sequence (“tag”) followed by several random nucleotides is tagged-PCR (T-PCR). In some cases, the last few bases at the 3’ end are also specified, such that the random sequence is flanked by defined sequences and thus resembles the primers used for DOP-PCR. Several low annealing temperature PCR cycles are performed, during which the tagged-primer anneals at random, copying the entire genome and generating fragments with complementary sequences to the “tag” at their 3’ ends. A new primer of identical sequence to the tag is then added, allowing exponential amplification of all DNA fragments containing tag sequences at either end.

In contrast to DOP-PCR, T-PCR, and PEP, linker adapter PCR (LA-PCR) does not employ degenerate primers. Instead the sample DNA is digested using a frequent cutting restriction endonuclease. This creates a heterogeneous population of DNA fragments that vary in size. An adapter oligonucleotide, designed such that it will specifically anneal to the single-stranded overhangs created by restriction enzyme digestion, is then ligated to the 5’ and 3’ ends of each fragment. This produces fragments composed of genomic DNA flanked by adapters of defined sequence. A primer complementary to the adapter is then used to amplify all ligated fragments. DNA synthesis is initiated from adapters at both ends of each fragment and consequently there is an exponential increase of amplified DNA. This methodology has been used for the generation of chromosome specific libraries and paint probes. More recently, linker-adapter approaches have allowed comparative genomic hybridization (CGH) analysis of microdissected pieces of formalin-fixed tissue.
Another method of whole genome amplification that utilizes primers of defined sequence is *Alu*-PCR. This method employs primers complementary in sequence to *Alu* repeats, approximately 900,000 of which are interspersed throughout the human genome. By initiating DNA synthesis from each of these points, amplification of virtually the entire genome can be achieved. Many of the fragments generated will extend from one *Alu* repeat to the next. The second repeat can also serve as a site for primer annealing, and, if it is in the correct orientation, the intervening DNA can be amplified from both ends of the fragment. The exponential amplification that this provides is highly efficient, leading to a dramatic increase in fragments amplified in this way.

*Alu*-PCR has been effectively employed for the generation of probes from cloned DNA fragments and flow sorted chromosomes. An advantage of this approach is that *Alu* repeats are

---

**FIGURE 34.1** Degenerate oligonucleotide primed PCR (DOP-PCR). (A) The only reagent significantly different to those found in a typical PCR amplification is the primer. In this case a mixture of semidegenerate primers, with specified 5’ and 3’ sequences separated by 6 random nucleotides, is employed. (B) Low annealing temperatures are used during thermal cycling, permitting the degenerate primers to anneal at numerous sites throughout the sample genome. (C) Primers that have annealed to sites that have complementary sequence to their 3’ ends can serve as initiation points for DNA synthesis. Many such sites exist and, consequently, the genome is copied from many positions. (D) The fragments amplified in this way can also serve as templates for amplification in subsequent cycles. (E) After several rounds of amplification most of the fragments present will have complementary sequences to the semidegenerate primer at both ends. (F) If the annealing temperature of PCR is then increased the semidegenerate primer starts to behave in a more specific manner, annealing only to sequences of extensive complementarity and thus amplifying the products of earlier cycles in an exponential manner.
particular to humans. This feature has allowed chromosome specific probes to be produced from somatic cell hybrids that contain single human chromosomes (or chromosome fragments) on a rodent background \(^{18}\) and has been used for amplifying human DNA prior to cloning.

Not only do minute tissue samples constrain the number of genetic loci that can be analyzed, they also place limits on the amount of cytogenetic data that can be obtained. Problems arise because techniques for the generation of metaphase chromosomes are inefficient when applied to small samples and require living, dividing cells. Fluorescent in situ hybridization (FISH) analysis has overcome some of these problems as it can be applied to interphase nuclei as well as metaphase chromosomes. However, technical limitations prevent analysis of more than a handful of chromosomes per cell using this method. Furthermore, FISH and PCR are not readily compatible and, consequently, single cell analyses have usually had to focus on analysis of DNA sequence or chromosomes, but not both.

An alternative cytogenetic technique, which can be applied to samples regardless of their phase of the cell cycle, is comparative genomic hybridization (CGH). \(^{19}\) CGH is a DNA-based method that provides data regarding the copy number of every chromosome. The CGH procedure requires at least 100 ng of DNA and, consequently, WGA is necessary if it is to be applied to small DNA samples. Methods that allow CGH to be performed on single cells have recently been described. \(^{1,20}\) The amount of DNA produced by certain WGA methods is sufficient to allow not only CGH but also multiple independent PCR analyses. Thus, comprehensive cytogenetic testing and multiple molecular genetic analyses can be performed on the same isolated cell. \(^{6,21}\) We determined that DOP-PCR provided the best templates for this purpose and have continued to optimize this form of WGA to a level sufficient for extensive PCR analysis and CGH.

II. MATERIALS AND METHODS

A. Sample Preparation and Cell Lysis

Using a micromanipulator, single cells were washed through four 5 to 10 µl droplets of PBS/0.1% polyvinyl alcohol, transferred to microfuge tubes containing 2 µl of proteinase K (125 µg/ml) and 1 µl of sodium dodecyl sulfate (17 µM), and then overlaid with mineral oil. DNA was released by incubation for 60 min at 37°C, followed by 15 min at 95°C.

B. WGA Using Degenerate Oligonucleotide Primed PCR (DOP-PCR)

The sample DNA was amplified by DOP-PCR (modification of Wells et al.). \(^{6}\) Amplification took place in a 50 µl reaction volume. A variety of conditions, as outlined in Table 34.1, were attempted. Optimal conditions were as follows: 0.5 mM dNTPs; 3.0 µM DOP primer (CCGACTCGAGNNNNNATGTGG); \(^{7}\) 1X buffer 3 and 2.5 U Polymerase mixture (Expand Long Template PCR System, Roche). Thermal cycling conditions were as follows: 94°C for 4.5 min, 10 cycles of 95°C for 30 sec, 30°C for 1 min, a 1°C/sec ramp to 68°C, 68°C for 3 min, 25 cycles of 95°C for 30 sec, 56°C for 1 min, 68°C for 3 min, and finally 68°C for 8 min. After amplification was complete, a 5 µl aliquot of amplified DNA was transferred to a new PCR tube and retained for single-gene testing, while the rest of the DOP-PCR product was used for CGH.

C. Tagged PCR (T-PCR)

Amplification of lysed single cells took place in a 50 µl reaction volume that contained: 0.1 mM dNTPs, 1.7 µM TAG-random primer (GAGTACCGACTTAGCCNNNNNNNNNN), 1X Taq polymerase buffer, and 5 U Taq polymerase. Thermal cycling conditions were as follows: 1 cycle at 94°C for 4.5 min, 30 cycles of 96°C for 0.5 min, 30°C for 1 min, 40°C for 1 min, 72°C for 2 min, and finally 1 cycle at 72°C for 10 min. Excess dNTPs and primers were then removed from the reaction mixture by passage through a spin column (S-400, Pharmacia). Further amplification
then proceeded with the addition of 8 μl of PCR buffer, dNTPs (final concentration 0.1 mM), 0.32 nmol TAG primer (GAGTACGCGACTTAGCC), and 5 U of Taq polymerase in a final volume of 80 μl. This mixture was heated to 94°C for 4.5 min and then subjected to 30 cycles of 96°C for 30 sec, 40°C for 1 min, 50°C for 1 min, and 72°C for 2 min. Finally the mixture was heated to 72°C for 10 min.

D. PRIMER EXTENSION PREAMPLIFICATION (PEP)

PEP was performed in a volume of 60 μl consisting of 6 μl of PCR buffer, 0.1 mM dNTPs, 33.3 μM random primer (N)₃₅, and 5 U Taq polymerase. Reactions were subjected to 50 cycles of 92°C for 1 min, 37°C for 2 min, a temperature ramp of 0.1°C/sec to 55°C, and then 4 min at 55°C.

E. WGA USING ALU-PCR

Lysed cells were amplified in 50 μl containing: 5 μl of PCR buffer, 0.2 mM dNTPs, 0.7 nmol primer1 (TCCCCAAGTGCTGGGATTACAG), 0.7 nmol primer2 (CTGCACTCCAGCCTGGG), and 2.5 U Taq polymerase. The amplification reaction was heated to 96°C for 5 min and then cycled according to the following conditions: 96°C 1 min, 40°C for 30 sec, 72°C 6 min, 50 times. The mixture was then given one final extension at 72°C for 10 min.

Stringent precautions against contamination, as discussed by Wells and Sherlock,²² were observed throughout cell isolation, lysis, and amplification procedures. Negative controls, composed of 2 μl of PBS taken from the final droplet used for washing each cell, were subjected to WGA. The amplification of a specific gene was then attempted using an aliquot of the resultant product. Successful amplification was indicative of contamination. Some WGA products from the control were also used for CGH. In this case any fluorescence detected on the chromosomes was indicative of contamination.

F. ASSESSMENT AND ANALYSIS OF WGA PRODUCTS

Various methods were employed to assess the different DOP-PCR methods. The range of DNA fragment sizes produced was determined by electrophoresis on a 1% agarose gel stained with ethidium bromide. Electrophoresis also revealed whether or not any of the degenerate or semidegenerate primers used had preferential annealing sites within the genome, indicated by distinct bands on the gel. The intensity of fluorescence observed on the gels was an indication of the amount of amplification. DNA concentration was also assessed by spectrophotometry, although this was found to have limited value.

---

**TABLE 34.1**

<table>
<thead>
<tr>
<th>Reagent/Condition</th>
<th>Variants Tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>dNTP (mM)</td>
<td>0.5 0.75 1.0</td>
</tr>
<tr>
<td>DOP-primer (μM)</td>
<td>0.3 0.5 1.0 3.0 5.0</td>
</tr>
<tr>
<td>Enzyme (U)⁠</td>
<td>1.5 2.0 2.5</td>
</tr>
<tr>
<td>Extension time (min)</td>
<td>3 5 7</td>
</tr>
<tr>
<td>Extension temp (°C)</td>
<td>68 72</td>
</tr>
<tr>
<td>MgCl₂ (mM)</td>
<td>1.5–8.0⁠</td>
</tr>
<tr>
<td>Buffer (number)</td>
<td>1 2 3</td>
</tr>
</tbody>
</table>

¹ Enzymes were supplied by Ambion (SuperTaq Plus) or Roche (Extend Long Template PCR System).
²⁵ Increments of 0.5 mM MgCl₂ were assessed.
³ Variant PCR buffers accompanying Roche ELT kit.
To estimate how many genetic loci were successfully amplified during WGA, we attempted to amplify multiple DNA fragments from 4 μl aliquots of WGA products. The loci amplified were FACC, D21S11, D21S1414, CFTR, D18S535, β-globin, hMSH2, APC, and amelogenin X and Y chromosome specific sequences.5

Preamplification by WGA involves additional PCR cycles and thus increases the likelihood that errors are introduced into the DNA sequence. To assess this potential problem, we amplified regions of the APC gene from aliquots of PEP and DOP-PCR products and then subjected them to mutation analysis or DNA sequencing. Mutation analysis took the form of restriction enzyme digestion or single strand conformational polymorphism (SSCP) as described previously.23 DNA sequencing of the APC gene was performed using the BigDye Terminator cycle sequencing kit (Applied Biosystems) according to the manufacturer’s instructions.

The fidelity of replication was also assessed for microsatellite loci (D21S11, D21S1414, D18S535). Microsatellites were amplified from 4 μl WGA aliquots using fluorescent-labeled primers. The resultant products were then analyzed using an ABI Prism 310 genetic analyzer, a capillary based electrophoresis, and DNA sequencing apparatus. Electrophoresis was performed as described elsewhere.24

**G. CHROMOSOMAL ANALYSIS USING COMPARATIVE GENOMIC HYBRIDIZATION**

WGA products were precipitated and labeled with fluorescence by nick translation. The amplified DNA from single cells was labeled with Spectrum Green-dUTP (Vysis) while DNA from a chromosomally normal individual (usually a male-46, XY) was labeled with Spectrum Red-dUTP (Vysis). Both labeled DNAs were precipitated together with 30 μg of Cot1 DNA. Precipitated DNA was resuspended in a hybridization mixture composed of 50% formamide, 2X SSC, 10% dextran sulphate, denatured at 75°C for 10 min, and then allowed to cool at room temperature for 2 min, before being applied to denatured normal chromosome spreads.

Metaphase spreads were usually derived from a normal male (Vysis) and were dehydrated through an alcohol series (70, 85, 100% ethanol for 3 min each) and air-dried. The slides were then denatured in 70% formamide/2 X SSC at 75°C for 5 min (20 X saline sodium citrate is 3 M NaCl, 0.3 M sodium citrate, pH 7). Following this incubation, the slides were put through an alcohol series at –20°C and then dried. The labeled DNA probe was added to the slides and a coverslip was placed over the hybridization area and sealed with rubber cement. Slides were then incubated in a humidified chamber at 37°C for ~72 h. After hybridization the slides were washed sequentially in 2 X SSC (73°C), 4 X SSC (37°C), 4 X SSC + 0.1% Triton-X (37°C), 4 X SSC (37°C), and 2 X SSC (room temperature), each wash lasting 5 min (modification of Levy, B. et al.).25 The slides were then dipped in distilled water, passed through another alcohol series, dried, and finally mounted in antifade medium (DAPI II, Vysis) containing diamidophenylindole to counterstain the chromosomes and nuclei.

Fluorescent microscopic analysis allowed the amount of hybridized DNA from amplified single cells (labeled green) to be compared to amplified DNA from the chromosomally normal control (labeled red). Computer software (Applied Imaging) revealed the red:green ratio along the length of each chromosome; deviations from a 1:1 ratio were indicative of loss or gain of chromosomal material.

**III. RESULTS**

**A. FRAGMENT SIZE AND AMPLIFICATION BIAS**

Electrophoretic analysis of DOP-PCR amplified DNA produced a smear, indicating that multiple distinct fragments of differing sizes were present. Using a commercially available mixture of Taq polymerase and Pwo or Tgo polymerase designed for amplification of long DNA templates, the
fragments were seen to range in size from 500 bp to 2 kb, the average fragment size being ~1300 bp. However, if the duration of the extension phase of the PCR was increased to 7 min, fragments in excess of 8 kb could be detected and the average fragment size increased to 3.5 kb. Other DOP-PCR methods that achieve amplification of fragments in excess of 10 kb have also been reported.26

Electrophoretic analysis also revealed evidence that DOP-PCR amplification is not entirely random, preferentially amplifying certain fragments of genomic DNA. Preferential amplification was revealed by the presence of distinct bands visible within the smear of DNA fragments (Figure 34.2). Similar results were observed using Alu-PCR and T-PCR, but PEP amplification of single cells did not produce sufficient DNA for visualization on an agarose gel. Preferentially amplified fragments were less apparent in amplifications that used the Taq/Pwo polymerase mixture.

B. COVERAGE OF THE TEMPLATE GENOME AND OVERALL YIELDS

The proportion of the genome efficiently amplified using different WGA methods was assessed by amplifying specific loci from aliquots of WGA products. Successful amplification of a given locus was indicative of its representation in the WGA product. Ten loci from different regions of the genome were tested. Approximately 90% were successfully amplified and were therefore present in the DOP-PCR and PEP products. This compares favorably with methods such as Alu-PCR (59% amplification success rate) and T-PCR (76% success rate). The amplification success rates of individual loci displayed some variation, ranging from 80% to 100% for DOP-PCR and PEP, suggesting that there are small differences in the efficiency with which certain genomic regions are amplified using these methods. Methods such as Alu-PCR and T-PCR displayed a much wider range of variation. In the case of Alu-PCR, the 19 to 86% variation in amplification rates that we observed is probably explained by the significant differences in the concentration of Alu repeats in different regions of the genome.
The amount of DNA produced by DOP-PCR far exceeds that generated using PEP and is only rivaled by other exponential amplification methods, such as T-PCR and Alu-PCR. Unfortunately, the precise quantity of DNA produced cannot be readily assessed as large numbers of PCR artefacts are generated during amplification and confuse spectrophotometric evaluation. These artefacts are of various sizes and cannot be easily removed. However, the fragments produced as a result of primer–primer interactions and other artefacts do not seem to interfere with subsequent PCR amplification of specific loci and do not hybridize to chromosomes if used for comparative genomic hybridization (CGH).

C. FIDELITY OF WGA

Several regions of the APC gene were amplified from aliquots of DOP-PCR or PEP products and their sequences assessed using various forms of mutation analysis or DNA sequencing. The use of single-strand conformational polymorphism (SSCP) analysis or restriction digestion confirmed the expected genotypes in all cells tested (100/100), with no alterations in DNA sequence detected. Similarly sequencing of amplified fragments provided the expected DNA sequence at each nucleotide position in 100% (20/20) of cases.

Although the amplification of unique DNA sequences from single cells is accomplished with high fidelity, the replication of repetitive loci, such as the microsatellites often used in linkage analysis or for fingerprinting, is error prone. The errors take the form of alleles that differ in size from that expected by a number of repeat units. These artefacts are presumably caused by polymerase slippage during the replication of repetitive loci. Although this is a significant limitation of WGA, preventing the use of DNA fingerprinting or quantitative PCR for aneuploidy detection, it is only seen when exceptionally small samples (e.g., single cells) are amplified.

D. CYTOGENETIC DATA FROM WGA PRODUCTS

We amplified the DNA of single cells using various whole genome amplification techniques and attempted to use the resultant products for CGH. The amount of DNA produced by PEP was found to be insufficient for CGH, while methods such as Alu-PCR and T-PCR amplified certain areas of the genome more efficiently than others, leading to uneven hybridization and poor aneuploidy detection. Our best CGH results were achieved following DOP-PCR amplification. Initial experiments were performed using cells derived from aneuploid prenatal samples. The use of DOP-PCR and CGH allowed trisomy for chromosomes 13, 14, 18, and 21 to be detected in single amniocytes. Subsequent experiments were able to demonstrate chromosome imbalance in various other types of sample, including microdissected tumor samples, oocytes, polar bodies, and cells isolated from preimplantation embryos.

IV. DISCUSSION

The advent of DNA amplification via the polymerase chain reaction has opened up the possibility of analyzing the DNA sequence of minute tissue samples, including single cells. However, routine PCR methods only permit the amplification of a small number of loci in each reaction. WGA methods have done much to overcome this limitation by providing enough DNA for multiple individual analyses to be performed. In the case of DOP-PCR, the genome of a single cell can be amplified to a sufficient level to allow more than 90 subsequent PCRs to be performed.

The proportion of the genome amplified by WGA depends on the particular method used and the amount of DNA template added to initiate the reaction. In our hands DOP-PCR and PEP provide the greatest genomic coverage with ~90% of loci successfully amplified from single cells. It is likely that further optimization, including the use of DNA polymerase mixtures that are designed for the amplification of long templates, will improve coverage further. If larger
Whole Genome Amplification from Single Cells and Minute DNA Samples

amounts of DNA are amplified (i.e., several cells) then almost all loci can be detected in the final WGA product.

Our data show that in most cases the amplification of coding sequences using DOP-PCR or PEP is achieved with high fidelity (Piyamongkol and Wells, unpublished), producing templates that are suitable for genotyping and mutation analysis. Similar results have also been reported using linker-adapter mediated PCR. These tests have allowed multiple genetic diagnoses to be performed on single cells with high accuracy. However, in the case of highly repetitive sequences, such as microsatellites, fidelity is poor with numerous artefacts frequently observed. This problem appears to be unique to WGA of single cells. It is not observed if repetitive sequences are amplified directly from single cells (i.e., without WGA), nor is it seen if WGA is performed on larger DNA samples. The artefacts produced take the form of DNA fragments that differ in size from that expected, having a number of repeat motifs added or deleted. These errors are much more pronounced and unpredictable than the “stutter” artifact commonly observed when microsatellites are amplified, although both phenomena are probably caused by factors such as DNA polymerase slippage, template switching, and strand displacement. The errors have confounded attempts to genotype microsatellites from single-cell WGA products, but cause no problems for WGA products made from larger numbers of cells. It is possible that the artefacts observed might be reduced in frequency or eliminated by using a different type of polymerase.

Some WGA techniques, such as DOP-PCR, generate sufficient DNA to perform tests that require large quantities of DNA. One example of such a test is comparative genomic hybridization (CGH), a molecular cytogenetic method that reveals losses and gains of chromosomal material throughout the genome. It is difficult to apply conventional cytogenetic techniques to small numbers of cells, as metaphase chromosomes are required. Most cells sampled are found to be in interphase regardless of the tissue examined. Even when cells in metaphase are acquired, it is difficult to ensure that a fully analyzable chromosome spread will be obtained after fixation and spreading. As a consequence of these problems, most cytogenetic analyses of single cells have employed fluorescent in situ hybridization (FISH).

Fluorescent in situ hybridization employs labeled DNA probes that hybridize to specific chromosomal regions, producing a discrete signal and thus revealing the copy number of individual chromosomes. Signals can be visualized in interphase as well as metaphase. However, it is only possible to accurately test a small number of chromosomes per cell. CGH, on the other hand, reveals the copy number of every chromosomal region over 5 to 10 Mb in size, but requires in excess of 100 ng of DNA to be successful. To perform single cell CGH, it is first necessary to achieve a 10,000-fold amplification of the genome. This has been accomplished using DOP-PCR and has led to some interesting scientific data concerning chromosome abnormality in early human embryos, oocytes, polar bodies, and microdissected tumor samples.

Most genetic studies of small tissue samples or single cells have been forced to focus on either DNA sequence or chromosomes because of the incompatibility of PCR and FISH techniques. Attempts to combine PCR and FISH have achieved only limited success, with reduced amplification efficiency and increased hybridization failure rates reported. The use of WGA can potentially overcome this limitation by generating sufficient DNA to allow CGH and also multiple PCR amplifications to be performed on the same isolated cell. We have successfully screened a DOP-PCR product generated from a single polar body for chromosomal imbalance (using CGH) and cystic fibrosis mutation (via PCR). It is clear that WGA techniques are powerful tools for analyzing small tissue samples and single cells in particular. The existing techniques have succeeded in expanding the number of genetic investigations that can be performed on minute tissue samples or unique and irreplaceable DNA samples. New WGA methods will undoubtedly extend the possibilities for research further still. Recent developments that show promise include new linker adapter mediated PCR (LA-PCR) methods, modifications of PEP (I-PEP), and protocols based on the principle of rolling circle replication.
The use of WGA will be an essential part of any attempts to assess aneuploidy or mutation in single cells using microarray technology. As with CGH, such tests require significant quantities of DNA. It is possible that certain WGA methods might also be adapted for the amplification of cDNA prior to microarray analysis of gene expression. However, amplification techniques that introduce promoter sequences to the ends of cDNA fragments and allow further rounds of transcription are likely to be more effective in this regard. With the burgeoning interest in noninvasive forms of prenatal genetic diagnosis, PGD, forensics, and cancer detection, the desire for detailed investigation of small DNA samples will inevitably increase further. It is likely that whole genome amplification methods will play an increasingly important role in these areas of study.

REFERENCES


I. INTRODUCTION

Single-cell PCR (polymerase chain reaction) is a powerful method for determining the genetic properties of an individual cell. However, since only one copy of the genome is present in a single cell, the number of genetic analyses that can be performed is significantly limited. Therefore, strategies have been developed to amplify the whole genomic content of single or small pools of cells prior to the analysis of specific loci. One of these methods is called primer extension preamplification (PEP). It increases the amount of template DNA by unspecifically amplifying the whole genome using random primers in a PCR reaction. Since its inception, the PEP technique has been used for the analysis of single spermatozoa, single or few blastomeres, and tumor cells (as well as fetal cells which have been enriched from the maternal circulation) and in conjunction with comparative genomic hybridization (see Chapter 34 by Dagan Wells and Mercedes Garcia Bermudez). However, whole genome amplification by PEP is not without pitfalls, the most important being allelic loss. Consequently, this method is rarely used for clinical applications.

A major focus of our research is the development of new risk-free prenatal diagnosis methods using fetal cells isolated from the maternal circulation. It is, therefore, essential that we are able to reliably analyze multiple loci in single cells. Our previous experience has shown that current PEP procedures do not lead to equal amplification of the entire genome and can actually lead to
loss of alleles differing by as little as a single nucleotide. For this reason, we have performed several extensive analyses of whole genome amplification methods. In this report, we provide a summary of our most pertinent observations regarding three different PEP procedures and a detailed description of our most effective protocol.

As we wanted to determine which PEP procedure was the most effective, we have made extensive use of real-time quantitative (TaqMan®) PCR assays, in order to accurately determine the extent of whole genome amplification process. The reason for using this technology is that it permits an accurate and reproducible quantitation of gene copy number. Furthermore, by being based on a 96-well system, it provides very fast and high throughput. A further advantage is that since the samples are analyzed directly in real time during the PCR amplification, this process does not require any post-PCR sample handling, thereby preventing contamination of subsequent assays with previously amplified PCR products. This is essential when dealing with small template quantities. In our examinations we amplified two genetic loci which we have extensively used before in other investigations: the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene on chromosome 12 and SRY gene on the Y chromosome. These previous studies indicated that these target loci are amplified with equal efficiency over a very broad range of DNA concentrations. The use of these two genes would also indicate if any difference in the amplification of autosomal or sex chromosomes occurred during the PEP process. Although not a prerequisite for analytical studies, in our experience this type of approach can considerably aid in the optimization of PEP protocols, as well as in the rapid assessment of troubleshooting strategies.

For the reliable and contamination-free isolation of single cells we used a PALM MicroBeam microdissection/laser pressure catapulting (LMM/LPC) system. Again, this is not an absolute requirement, as we have on previous occasions been able to reproducibly and accurately isolate single cells using a manual micromanipulator and a finely drawn glass capillary. A comparison of different microdissection techniques can be found in Chapter 4.

II. EXAMPLE

A. PREPARATION SLIDES

The mononucleated cells from male peripheral blood samples were separated with a single density gradient using Ficol-Paque-1119® (Sigma, St Louis, MO). The cells were resuspended in PBS containing 1% bovine serum albumin (BSA) (Sigma) and transferred to the glass slides, which were covered by a 1.35 μm thin polyethylene membrane (PALM Microlaser Technology, Munich, Germany). The supporting membrane facilitates subsequent laser catapulting. The cells were fixed with methanol and stained with May-Grünwald-Giemsa (Sigma), a procedure that we routinely use in our laboratory. A simple, cost-effective way to obtain single cells is to prepare blood smears on normal glass slides. From these, individual cells can be removed by mechanical micromanipulation using a finely drawn glass capillary. Alternatively, smears on membrane coated slides were used for subsequent laser microdissection.

B. MICROMANIPULATION OF SINGLE CELLS USING LMM/LPC

The LMM/LPC system (PALM Robot-MicroBeam, PALM Microlaser Technology, Munich, Germany), marketed by PALM Microlaser Technology, uses a 337-nm high quality pulsed nitrogen laser, which can be used to microdissect selected cell clusters or single cells (LMM or Laser Mediated Microdissection). These single cells or clusters can then be catapulted directly into the cap of a routine PCR tube (located conveniently above the glass microscope slide) by the subsequent application of a high energy pulse of the highly focused laser beam. This procedure is termed Laser Pressure Catapulting (LPC). After this step, the laser-catapulted single cells were spun down into the reaction vessel containing 20 μl of the PEP-PCR mixture and stored at –20°C. Prior to PCR
analysis the tubes were again centrifuged at 13,000 g for 20 min in a table top centrifuge. As discussed above, we have determined the PALM laser microdissection system to be very effective for the reliable and rapid isolation of single cells. We have, however, on numerous previous occasions obtained very good results with a considerably less sophisticated manual micromanipulation device, using a finely drawn glass microcapillary.

C. PEP-PCR Protocol

We had previously obtained unsatisfactory results using the original PEP protocol described by Zhang and colleagues, as well as a slight modification thereof in which the concentration of MgCl₂ was increased to 5 mM. For this reason, we have recently examined three different PEP conditions (Table 35.1), which we tested on single cells isolated by laser microdissection. As a control we used total genomic DNA (<10 ng) isolated from peripheral blood. These different PEP protocols chiefly differed in the type of Taq polymerase employed or in the composition of the PCR buffer. In this manner, the mixture of PEP I contained AmpliTaq-gold polymerase in the standard TaqMan Universal Master Mix (Applied Biosystems), whereas for PEP II we used AmpliTaq-gold polymerase in a K⁺ free PCR buffer. For PEP III we used the Expand Long Template (ELT) PCR system (Roche Diagnostics GmbH). In essence, the same amplification steps were used for each of the three PEP procedures (Table 35.1). However, the PEP protocols I and II were preceded by a digestion with uracil-N-glycosylase for 2 min at 50°C. This enzyme was then deactivated by heating to 95°C for 10 min, which also served to activate the AmpliTaq Gold. All three procedures incorporated slow ramping by 0.1°C/min from 37°C to 55°C (Table 35.1). In general, 6 µl of the PEP treated material was used for the subsequent real-time PCR assay.

D. Analysis of PEP Efficiency Using Real-Time Quantitative PCR

Although many studies have reported on the use of PEP procedures, very few have attempted to address the efficacy of this method. Previously we have reported that the PEP procedure results in unequal amplification and may even lead to allelic loss. In our previous studies we used two different tools to examine PEP efficacy: direct sequencing of heterozygous loci differing by a single nucleotide and fluorescent PCR analysis of highly polymorphic microsatellite markers. Although very

| Table 35.1 | Summary of the Three PEP Protocols |
|---|---|---|
| Buffer | TaqMan Universal Master Mix | K⁺ free PCR Buffer | Roche ELT buffer III |
| dNTPs | TaqMan Universal Master Mix | 250 µM each | 250 µM each |
| 700 µM N15 | 2 µl | 2 µl | 2 µl |
| Enzyme | 0.75 U AmpliTaq Gold | 0.75 U AmpliTaq Gold | 3.75 U ELT |
| Volume | 20 µl | 20 µl | 20 µl |
| 50°C for 2 min | + | + | + |
| 95°C for 10 min | + | + | + |
| 95°C for 1 min | + | + | + |
| 37°C for 1 min | + | + | + |
| ramping step | + | + | + |
| 55°C for 4 min | + | + | + |
| 60°C for 1 min | + | + | + |
| 72°C for 1 min | + | + | + |

* Ramping step was 0.1°C/min from 37°C to 55°C.
precise, these assays are tedious to perform and are not well suited to monitor a series of experimental procedures. For this reason we established a single-cell TaqMan real-time PCR assay in order to determine the whole genome amplification efficiency of the three different PEP procedures tested. Briefly, this real-time assay monitors PCR product accumulation by the specific release of a fluorescent marker during each cycle of the reaction (the chapters by Stephen A. Bustin and Houng L. Vu et al. contain a more detailed discussion of this and other quantitative PCR methods). The amount of input template can be determined from the point during the exponential phase of amplification where the accumulation in fluorescence intensity crosses a defined threshold. This point is referred to as the threshold cycle ($C_T$). Consequently, the higher the amount of input template DNA, the lower the numbers of cycles required for crossing this threshold value, and hence the lower the $C_T$ value will be. In our studies, we used the $C_T$ value to estimate the efficiency of PEP procedure.

To reduce the degree of variability due to extraneous factors such as pipetting errors and unequal amplification between cycler wells, we used a multiplex TaqMan assay for the simultaneous detection of GAPDH and SRY genes. We have extensively tested this system previously and have obtained highly reliable quantitative results on small template copies. All assays were performed using an ABI PRISM 7000 Sequence Detector (Applied Biosystems), analyzing all samples in triplicate.

Although it may be possible to use other quantitative PCR systems for this type of analysis, we do not have extensive experience on their use and cannot comment on their efficacy. We have, however, ascertained that the ABI PRISM 7000 or 7700 TaqMan systems are highly robust and, under appropriate conditions with stringent control parameters, can detect less than twofold differences in the amount of starting template. TaqMan real-time PCR technology is not essential for studying the efficacy of a particular PEP procedure or for establishing PEP-based protocols in the laboratory. We have, however, found it to be useful for providing a very rapid and precise evaluation of the degree of amplification offered by a particular PEP procedure, and it is a considerable aid in troubleshooting.

E. INTERPRETATION OF THE RESULTS

In our examination we determined that the PEP III procedure was the most effective preamplification procedure. This is evident from Figure 35.1, where the amount of efficacy of each PEP procedure was analyzed by real-time PCR for the GAPDH gene. This analysis indicated that PEP III had a $C_T$ value of 29, while PEP II had a $C_T$ value of 31, and PEP I had a $C_T$ value of approximately 33. Assuming that each cycle leads to a doubling of the PCR product, then the PEP II procedure was about 25% ($\Delta C_T = -2; 2^{-2} = 1/4 = 25\%$) as effective as PEP III, while PEP I only lead to about $1/16th$ ($\Delta C_T = -4; 2^{-4} = 1/16$) of the fold-amplification of PEP III (Figure 35.1).

The actual-fold amplification of the PEP III procedure is shown in Figure 35.2, where we compared the difference in the amount of GAPDH template between a sample of untreated genomic DNA, and one that has been subjected to PEP III. The untreated sample has a $C_T$ value of 25 whereas the PEP treated sample has a $C_T$ value of approximately 20. Again, assuming that each cycle leads to a doubling of the PCR product, this indicates that the PEP treatment has amplified the amount of input template 32-fold ($\Delta C_T = 5; 2^{5CT} = 32$). Furthermore, as we were only using 6 µl out of a total of 20 µl, the dilution factor implies that the PEP amplification was up to approximately 100-fold.

By examining two different loci — the SRY gene on the Y chromosome and the GAPDH gene on chromosome 12 — simultaneously by multiplex real-time PCR, we could determine whether both these target genes had been amplified with equal efficiency. Our studies indicated that the PEP III procedure led to an unequal amplification of these two target loci, as in several instances either the SRY gene or the GAPDH gene was preferentially amplified. This feature was found to be attributable to the PEP procedure and not due to an artifact of the TaqMan PCR assay, as no such feature was observed when examining single cells which had not been subjected to PEP, in
Primer Extension Preamplification (PEP) of Single Cells: Efficiency and Bias

FIGURE 35.1 Efficacy of three different PEP procedures as analyzed by real-time PCR for the GAPDH gene. The lower the number of PCR cycles required for the target template to traverse the threshold limit (\(C_T\) value), the more abundant the amount of input template. Therefore, the PEP III procedure results in the most optimal amplification, as it had the lowest \(C_T\) value (\(C_T = 29\)), compared to PEP II (\(C_T = 31\)) and PEP I (\(C_T = 33\)).

FIGURE 35.2 Degree of amplification afforded by the PEP III procedure. The difference between the number of PCR cycles required to obtain the threshold value (\(C_T\) value) between the PEP III treated material and non-PEP treated material indicates up to 32-fold amplification (\(\Delta C_T = 5; 2^5 = 32\)) of input template. Since only a third of the PEP material was used, this result indicates that the total amplification by PEP could be close to 100-fold.
which case no significant difference in the respective C<sub>T</sub> values was noted. Indeed, we found this phenomenon of unequal amplification by PEP to occur even when using small amounts (<10 ng) of total genomic DNA. Of note, the PEP procedure did not appear to alter the sensitivity of the single-cell PCR analysis for the two loci interrogated. Both were detected with approximately 50% of the cells examined by singleplex PCR, or about 33% of the cases for both genes when using multiplex PCR, irrespective of whether the target cell had been treated by PEP or not (Table 35.2).

F. Pitfalls and Troubleshooting Tips

PEP does not amplify the genome evenly; we and others have noticed that certain alleles are sometimes completely lost. This is a major concern for diagnostic purposes. For instance, in our previous studies we had focused on mutations of the β-globin gene, where we determined that the results of the PCR analysis following PEP would indicate that the tested cell was either homozygous affected or normal but not heterozygous for the mutation interrogated. Obviously, the determination that a fetus may be homozygous affected will have dire consequences in a prenatal diagnostic setting. Our studies also indicated that the allelic loss was not restricted to point mutations, but that microsatellite markers may be affected in a similar manner. As we have previously observed that this affect can influence the analysis of several adjacent markers, it appears that allelic loss occurs over fairly large regions of the chromosome being analyzed. PEP therefore may not only lead to the unequal amplification of two different loci (located on the same or different chromosomes) but PEP may also lead to the loss of an allele of a particular gene, or a larger region surrounding the gene being interrogated. It is, therefore, important that appropriate strategies be adopted to avoid being misled by such PEP artifacts. This can be achieved by examining two closely linked markers, e.g., the mutation in question and a linked SNP (single nucleotide polymorphism). If the closely linked SNP is present, the chance that the mutant allele has been lost by PEP may be assumed to be quite low.

As the PCR assays we have previously used to test the efficacy and reproducibility of the PEP procedure are tedious, and as we wanted to determine which PEP procedure resulted in the most optimal random amplification, we have made extensive use of a real-time TaqMan PCR. The advantage of this system lies in its ability to examine numerous samples at the same time. By using a multiplex assay for simultaneous quantification of GAPDH and SRY genes, we were able to rapidly determine whether one loci was preferentially amplified over the other by PEP. That such artifacts can occur following PEP treatment is evident from our examinations on single cells and small amounts of genomic DNA (not shown) amplified by PEP. We also determined that in the case of GAPDH or SRY, this effect was random and not locus specific. PEP treatment did not noticeably alter the sensitivity of the single-cell PCR analysis. We were able to detect either the GAPDH or SRY loci in approximately 50% of the single cells analyzed, and both loci in about 32% of single cells analyzed regardless of whether the template had been treated by PEP or not (Table 35.2).

<table>
<thead>
<tr>
<th>TABLE 35.2</th>
<th>Detection of Single (SRY of GAPD) or Multiple Loci (Both SRY and GAPDH) in Single Cells either Treated by PEP or Not Preamplified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GAPDH</td>
</tr>
<tr>
<td>PEP-TaqMan</td>
<td>85/164</td>
</tr>
<tr>
<td>(164 single cells)</td>
<td>(51.8%)</td>
</tr>
<tr>
<td>TaqMan</td>
<td>47/96</td>
</tr>
<tr>
<td>(96 single cells)</td>
<td>(48.9%)</td>
</tr>
</tbody>
</table>

www.taq.ir
IV. DISCUSSION

The ability to reliably examine the genetic content of a single cell is pursued by scientists working in a diverse range of disciplines. Clinically, the analysis of single cells has made the biggest impact in the field of prenatal diagnosis in the form of preimplantation genetics. Since the number of PCR analyses that can be performed on a single cell is limited by its single genome content, several strategies such as PEP have been developed which attempt to amplify this DNA in order to facilitate numerous analyses. That such procedures can be very effective is illustrated by the method presented, which might lead to an almost 30-fold amplification of the genomic content of a single cell and almost 100-fold that of limiting amounts of genomic DNA.

Our studies, however, indicate that this amplification process can also lead to errors, even to the actual loss of one of the alleles at a locus. In a prenatal diagnostic setting, such an error cannot be afforded as it may determine decisions regarding the termination of a pregnancy. Even when no actual loss occurs, not all loci or even alleles of a particular gene might be amplified with equal efficiency. Therefore, material that has been subjected to PEP treatment is unsuitable for subsequent quantitative analysis, such as the fluorescent PCR analysis of microsatellite markers for the determination of chromosomal ploidy, gene duplication, or loss.

In this study we have used advanced technology to assay the efficacy of the PEP method, namely, a laser-assisted microdissection tool and real-time PCR for the analysis. These are surely not necessary for the majority of experiments. Retrieval of single cells can also be effectively achieved using traditional mechanical micromanipulation systems, while normal PCR assays can be used for the monitoring of the PEP procedure. However, such analyses are important in order not to be mislead by PEP artifacts. We would, nevertheless, recommend the use of multiplex analyses (to minimize the number of handling steps), perhaps in combination with fluorescent detection of PCR products (to increase sensitivity). In addition, one should use small amounts of total genomic DNA (<10 ng) when starting to set up a PEP protocol.

Strict precautions should be taken to avoid any potential sources of contamination, including the use of separate rooms for single-cell retrieval, PCR/PEP procedures and subsequent analysis. We also recommend the use of N-uracil glycosylase systems to prevent carryover and filter-blocked pipette tips. Another important point to consider is the condition of the starting material. Our own investigations have shown that certain cells, such as mature erythroblasts, exhibit considerably higher rates of allele dropout than comparable lymphocytes retrieved from the same blood sample. This may be attributable to the apoptotic-like phenotype of many of these cells, including fragmented nuclear DNA as detected by the TUNEL assay.

It is to be expected that many of the current problems will be seriously reduced, if not completely eliminated, in the future. Should this be the case, then, it is possible that comparable advances in CHIP technology will permit the analysis of several thousand loci from a single cell, thereby opening an entire new realm of diagnostic opportunities.

ACKNOWLEDGMENTS

This work has been supported by the Swiss National Science Foundation (3200-059275.99/1) and the National Institute of Health (Contract Number N01-HD-4-3202). We would like to thank Anurag Kumar Gupta for comments on the manuscript.

REFERENCES

6. Hahn, S. et al., Allele drop-out can occur in alleles differing by a single nucleotide and is not alleviated by preamplification or minor template increments, *Genet. Test.*, 2, 351, 1998.
In Vitro Evolution
36 In Vitro Evolution through PCR-Mediated Mutagenesis and Recombination

Sarah B. Faulkner, Heather A. Syrett, Eric A. Davidson, and Andrew D. Ellington

CONTENTS

I. Introduction ................................................................................................................................. 371
   A. Mutagenic PCR ...................................................................................................................... 372
   B. In Vitro Recombination via DNA Shuffling ........................................................................ 372
II. Materials And Methods .............................................................................................................. 376
   A. Mutagenic PCR ...................................................................................................................... 376
      1. Mutagenesis by Polymerization in the Presence of Unbalanced dNTP Concentrations and Manganese ................................................................. 376
      2. Mutagenesis via the Incorporation of Modified Nucleotides ........................................ 376
   B. In Vitro Recombination via DNA Shuffling ........................................................................ 377
      1. Gene Fragmentation via DNase I Digestion ................................................................... 377
      2. Reassembly of DNA Fragments by PCR ....................................................................... 378
      3. Amplification of Fragment Libraries ............................................................................ 379
III. Results and Discussion ............................................................................................................ 379
Acknowledgments .......................................................................................................................... 379
References ....................................................................................................................................... 379

I. INTRODUCTION

Natural evolution occurs via two major mechanisms: random mutation followed by selection and recombination of the selected mutations. In vitro evolution has frequently proven to be an excellent tool for engineering functional nucleic acids and proteins with optimized or novel properties, and it should come as no surprise that laboratory techniques rely on the same basic mechanisms as does nature. However, because of the control which can be exercised over DNA replication in vitro, the number and type of mutations that can be introduced and recombined vastly exceeds the number typically seen during natural selection. The laboratory evolution of functional nucleic acids and proteins is correspondingly much faster than in nature. An excellent demonstration of this was the evolution of antibiotic resistance elements. While many of the same resistance mutations were found following "test tube" evolution, as had previously been observed in nature, the mutations arose much more quickly and were recombined into one highly resistant gene within just a few generations.
A. Mutagenic PCR

There are two basic methods for increasing mutational rates. Both methods take advantage of the relatively low fidelity of the *Thermus aquaticus* (Taq) DNA polymerase since this enzyme generates mutations at a rate of $1 - 2 \times 10^{-4}$ per residue per duplication during thermal cycling. In the first method, the fidelity of Taq DNA polymerase is further decreased by adding manganese ions in the presence of increased concentrations of magnesium, and skewing the ratios of the four natural nucleotides. This method yields 0.7% mutations per position per PCR. The other method relies on the introduction of synthetic nucleotide analogues such as (6-(2-deoxy-B-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one (P-2'-deoxyribose or dP) and 8-oxo-2'deoxyguanosine (8-oxodG), which can form ambiguous base-pairs with each of the four natural nucleotides (Figure 36.1). Once this “omni-base” has been initially misincorporated during template-directed polymerization, it will further facilitate the misincorporation of one or more “wrong” natural nucleotides at or near the corresponding position on the complementary strand. A variety of modified 2' deoxyribonucleoside triphosphates, including 8-oxo-2'-deoxyguanosine-5'-Triphosphate and 2'-deoxy-P-nucleoside-5'-Triphosphate, are available through TriLink Biotechnologies (Sorrento Mesa, CA).

B. In Vitro Recombination via DNA Shuffling

DNA shuffling, an *in vitro* recombination technique, was developed in 1994 by Willem Stemmer and has proven to be an extremely effective method for *in vitro* evolution. DNA shuffling has been

![Chemical structures of nucleotide analogues](https://www.taq.ir)

In Vitro Evolution through PCR-Mediated Mutagenesis and Recombination

used in the course of in vitro evolution experiments in which a variety of protein phenotypes were altered or optimized, including changes in substrate specificity, enhancing fluorescence, and improving thermal stability and expression. The DNA shuffling process is itself mutagenic, with a mutagenesis rate of 0.7% per position per PCR. This rate is comparable to that attained using mutagenic PCR as stated above. Thus, DNA shuffling on its own can be used to generate libraries for screening or selection. In instances where an even higher mutation rate is desired, mutagenic PCR can be combined with DNA shuffling.

Instead of generating sequence diversity, an experimenter can also shuffle extant sequence diversity by using a set of homologous alleles or phylogenetic variants as a starting point. Stemmer and his coworkers have touted such techniques as “family shuffling.” In an initial example, they shuffled naturally occurring homologues of four cephalosporinase genes that shared between 58 to 82% sequence identity, ultimately yielding a mutant with a 540-fold improvement in activity. Interestingly, DNA shuffling is not limited to single genes or even gene families; entire operons can also be shuffled using these techniques. For example, Maxygen researchers evolved the 2.3-kb arsenic resistance operon containing the genes arsR, arsB, and arsC over three rounds of shuffling and selection, and found variants that could grow in 0.5 M arsenate, a 40-fold increase in resistance. Unfortunately, the inherently mutagenic nature of shuffling can lead to the high frequency accumulation of detrimental point mutations in the shuffled genes. To obviate this problem, the Arnold lab has come up with several easily implemented measures to optimize DNA shuffling, including using DNA polymerases that have greater fidelity and using a plasmid rather than a PCR product as a template.

DNA shuffling in its most basic form consists of a three-step process (Figure 36.2). First, a given set of alleles or other genes that bear sequence similarities are digested into smaller fragments. This has traditionally been done with DNase I or more recently with restriction enzymes. The fragments are then randomly reassembled by overlap extension in the absence of gene-specific primers. The randomly recombinated genes are then amplified via the polymerase chain reaction using gene-specific primers. The result is a randomized gene library that in many instances will contain virtually all possible recombination products, including multiple recombination events involving different parental genes.

While the isolation of cleavage products is relatively simple and gives some control over the length of regions that will be recombined, other methods have also been worked out that achieve much the same results. For example, random-priming recombination (RPR) uses short random primers to generate mutagenized DNA fragments from the genes to be shuffled. Random sequence hexamers are annealed at low stringency to a gene or set of genes and are extended by DNA polymerase I. Incomplete homology between a random primer and a given gene inherently introduces point mutations in the gene products. Random-priming synthesis products are kept small by using high primer concentrations (resulting in multiple initiations or blocked extensions). The synthesis products are size-filtered so that template DNA and other large molecules are removed. The resultant randomized fragment library is then reassembled into a full-length gene library by allowing the fragments to prime one another’s extension. The gene library is ultimately amplified with outside primers containing (for example) BamHI and Ndel restriction sites. The full-length reassembly products are then digested with these restriction enzymes, cloned into a vector, and used for screening or selection. The difference between RPR and shuffling lies largely in the method of fragment production. RPR may have the added advantage of introducing a more predictable rate of point mutation during random-priming synthesis.

A second method, the Staggered Extension Process (StEP), relies on only PCR amplification to recombine genes. In StEP, a defined primer set is used and extension times are kept relatively short (30 sec for denaturation and a 5-sec annealing and extension at 55°C) so that short homologous extension products will accumulate and reprime synthesis on other templates (Figure 36.3). The extension process is continuously cycled until a full-length product is amplified; in the original paper an astounding 80 cycles.
Several other methods have recently been described that do not rely upon homologous recombination to generate products; these methods are particularly useful for generating variants of enzymes or functional nucleic acids that initially contain little or no sequence homology or similarity. In incremental truncation for the creation of hybrid enzymes (ITCHY), a method pioneered by the Benkovic lab, the genes to be shuffled are cloned into vectors and the vectors are cleaved to yield a potential junction between the two genes. The genes are digested with exonuclease III and S1 nuclease and the two vectors are then ligated together, creating a chimeric gene with a single recombination site. Exonuclease III has a catalytic activity that is salt concentration-dependent, and the extent of digestion can be finely controlled by using NaCl to slow base-pair excision to ≤10 bp/min. At short intervals, aliquots of the digestion reactions are removed and quenched by adding yet more salt and decreasing the pH. Following ligation of the aliquots, products can be size-selected to yield chimeras that contain a series of overlaps near a desired junction. Two out of three of these overlaps will be out of frame, but the method nonetheless allows the recombination of genes with little initial homology. For example, the *E. coli* and human glycaminde ribonucleotide transformylase genes were recombined to form functional chimeras, despite the fact

---

that these genes shared only 50% sequence homology. A variant of ITCHY, known appropriately as SCRATCHY, combines the technique with Stemmer’s original version of gene shuffling, allowing an even larger number of chimeras to be created.

Finally, nonhomologous random recombination (NRR) relies on the blunt-end ligation of multiple selected DNA fragments. Genes to be shuffled are first DNase I–digested, and then filled in with T4 DNA polymerase. The resulting set of fragments is then blunt-end ligated, generating...
chimeras that contain a variety of sequence-lengths, orientations, and orders that may be utterly different from the parental DNA molecules. Ultimately, adaptors are ligated to the end of the recombined fragments and the library is amplified. Because multiple, nonhomologous junctions are generated by this method, it is of little utility for protein engineering. However, Liu and his coworkers have used this method to generate novel antistreptavidin aptamers that present oligomeric binding sites to an oligomeric target.25

II. MATERIALS AND METHODS

A. MUTAGENIC PCR

1. Mutagenesis by Polymerization in the Presence of Unbalanced dNTP Concentrations and Manganese (Adapted from Reference 7)

0.4 fmol template DNA per μl total reaction volume
2.5 units *Taq* DNA polymerase for a typical reaction volume of 50 μl

Typical PCR components:
- 0.1 to 0.5 μM each PCR primer
- 0.2 mM each dNTP
- 60 mM Tris-HCl (pH 8.5)
- 15 mM NH₄SO₄
- 2 mM MgCl₂

Mutagenic Components:
- 0.8 mM additional dCTP and dTTP
- 4.8 mM additional MgCl₂
- 0.5 mM MnCl₂

A single mutagenic PCR is carried out over ~ 25 thermal cycles of 94°C for 30 sec, 67°C for 30 sec, and 72°C for 1 min/kb of desired product. Alternatively, a successful two-step PCR of 94°C for 30 sec and 72°C for 1 min/kb has been suggested by Matsumura et al.26 In the mutagenic reaction, manganese ions and increased magnesium ion concentrations decrease the fidelity of the enzyme, while the skewed nucleotide ratios further increase the probability that an incorrect base will be misincorporated by the polymerase.7 The rate of mutagenesis can be optimized by preparing a 10X stock of the mutagenic components, stored at –20°C, and adding varying amounts to obtain higher or lower mutation rates.8,26

Cadwell and Joyce report that this method yields a mutation rate of approximately $7 \times 10^{-3}$ per nucleotide per duplication (roughly equivalent to 0.7% per position per PCR).7 Low frequencies of insertions and deletions have been observed,7 and thus this method will tend to avoid frame shifts which could result in a loss of protein function. While this protocol has a strong bias toward transition mutations over transversion mutations,27 it has been successfully used for the directed evolution of proteins.28–30

2. Mutagenesis via the Incorporation of Modified Nucleotides (Adapted from Reference 9)

0.5 fmol template DNA per μl total reaction volume
5 Units of *Taq* DNA polymerase for a typical 50 μl reaction
Typical PCR components:
0.1 to 0.5 μM each primer
0.2 mM each dNTP
60 mM Tris-HCl (pH 8.3)
15 mM NH₄SO₄
2 mM MgCl₂

Mutagenic components:
0.2 mM each of dPTP and/or 8-oxo-dGTP

The target sequence is amplified in two stages. First, a PCR with modified nucleotides is carried out over 30 thermal cycles of 92°C for 1 min, 55°C for 1.5 min, and 72°C for 5 min. Alternatively, the authors have also employed a two-step PCR consisting of 94°C for 30 sec and 72°C for 1 to 2 min/kb for 30 to 35 cycles. Additional extension times may be necessary for the efficient incorporation of nucleotide analogues. The incorporation of modified nucleotides generates base mismatches since the modified nucleotides can form any of a variety of base-pairs (Figure 36.1). After the mutagenic PCR, a second, standard PCR (see text above) is carried out, using 1 μl of the mutated product as a template. The second amplification reaction serves to fix mutations due to misincorporation while simultaneously diluting the proportion of modified bases in the final product. A largely natural DNA product — containing few or no modified bases — can be important for subsequent manipulations such as cloning.

Modified nucleotide mutagenesis has reported mutation rates of 9.7 × 10⁻³ per residue after two cycles of PCR and 3.2 × 10⁻² per residue after 20 cycles of PCR. The desired level of mutation can be readily optimized by varying both, the ratio of normal dNTPs to analogs and the number of PCR cycles. All four transition mutations can be generated using this protocol due to the presence of dPTP, although A→G and T→C mutations arise almost four times more frequently than G→A and C→T mutations. Only 2 of the 8 transversion mutations are frequently generated (A→C and T→G mutations due to the 8-oxo-dGTP), although at a lower level than the transition mutations. Again, this method yields negligible numbers of insertions or deletions, and has been successfully used in a number of directed evolution experiments.

B. IN VITRO RECOMBINATION VIA DNA SHUFFLING

In the Ellington lab, DNA shuffling has generally been employed in its simplest form, as originally delineated by Stemmer et al., using the following protocol:

1. Gene Fragmentation via DNase I Digestion

In a total reaction volume of 100 μl add:
2 to 3 μg DNA
50 mM Tris-HCl, pH 7.4
1 mM MgCl₂
DNase I enzyme (see explanation below for amounts)
Water to 100 μl

The reaction conditions for a particular template must first be determined. It is recommended that a series of reactions with serial dilutions of DNase I and varied reaction times be carried out in order to identify optimal cleavage conditions. Increasing DNase I concentration and/or incubation times will decrease fragment sizes and therefore more recombination sites will be introduced into the reconstructed genes.
Incubate the above reaction components for 10 min at room temperature (~ 25 to 27°C), taking aliquots at convenient intervals and stopping the reaction by combining the aliquots with a 1:1 mixture of phenol:chloroform or EDTA. The size of the cleaved fragments can be conveniently monitored by a 2 to 5% LE agarose gel. Their relative amounts will depend on the desired size of the fragments. For example, 3 µl of 0.1 U/µl diluted DNase I incubated for 6 to 7 min with ~ 2 µg DNA will degrade a ~ 2 kb gene into 100 to 600 bp fragments. Fragments of a desired size could also be purified following agarose gel electrophoresis, allowing somewhat better control over fragment size and creating a more uniform pool of fragments. However, two things to consider when purifying gel fragments are the potential decrease in yield (see reassembly considerations) and the introduction of additional mutations by UV irradiation of the DNA.

Once optimal fragment sizes have been generated, extract the DNase I enzyme by mixing the reaction with 100 µl of a 1:1 mixture of phenol:chloroform. Vortex, centrifuge, remove the aqueous layer, mix with an equal volume of chloroform, and repeat the extraction (to ensure that all phenol has been removed). Precipitate the DNA with ethanol, then resuspend the fragments in a convenient volume.

**Ethanol precipitation:**

1. Add 2.5 volumes of 4°C 100% ethanol.
2. Add 1 µl of 20 mg/ml glycogen.
3. Invert several times or vortex to mix thoroughly.
4. Precipitate DNA at –80°C for 15 to 20 min or –20°C for 3 h or more.
5. Centrifuge at 10,000 rpm for 30 min.
6. Pour off the supernatant and wash the pellet with 100 µl of 70% ethanol (cold or room temperature).
7. Pipette off the 70% ethanol and air-dry the pellet (a SpeedVac or vacuum dessicator may also be used).

### 2. Reassembly of DNA Fragments by PCR

In a total reaction volume of 20 µl add:

- DNA fragments
- 0.2 mM each dNTP
- 60 mM Tris-HCl (pH 9.0)
- 15 mM NH₄SO₄
- 2 mM MgCl₂
- 1 Unit Taq DNA polymerase

A single reassembly without primers is carried out for ~ 40 thermal cycles of 94°C for 30 sec, 50 to 55°C for 30 sec, and 72°C for 30 sec. An initial “hot start” of 94°C for 1 min can be useful in avoiding low-fidelity priming events and promoting homologous recombination. Alternatively, a two-step reassembly program of 94°C for 30 sec and 72°C for 30 sec with a hot start of 94°C for 1 min can also be used. It should be noted that gene reassembly frequently must be optimized with respect to the amount of DNA fragments added and the precise thermal cycle employed. For example, a series of reassembly reactions can be carried out where the amount of DNA fragments added is increased in convenient intervals from 2 to 10 µl of the 20-µl total reaction volume. A set of PCR optimization buffers, such as those made by Invitrogen (Carlsbad, CA), may also aid in finding the ideal conditions for both the reassembly and the following amplification.
3. Amplification of Fragment Libraries

In a total reaction volume of 50 µl add:

- Diluted reassembly product
- 5 µl 10 X PCR buffer (same as that used above)
- 0.8 µM primers
- Sterile distilled H₂O to 50 µl

A single-PCR reassembly with primers is carried out over 15 to 20 thermal cycles of 94°C for 30 sec, 50°C for 30 sec, and 72°C for 45 sec, again using a hot start. Extended elongation times of 1 or more minutes per kb of DNA product have also been successfully used. Stemmer recommends a 1:40 dilution of the reassembly product into the final amplification, but this is also a variable for optimization. The number of cycles necessary for successful or sufficient amplification may vary. Typically, a 5-µl aliquot may be removed just before the end of the elongation cycle and analyzed by agarose gel electrophoresis to determine if additional cycles of amplification may be necessary.

III. RESULTS AND DISCUSSION

The mutagenized DNA pools derived by any of these methods can be cloned into a suitable vector, introduced into an appropriate organism, and screened or selected for function. For example, screening methods can vary from a technically simple visual screen to more complex fluorescence activated cell sorting (FACS) screens. Irrespective of the means by which genes are screened or selected, the degeneracy and integrity of the initial library should always be checked by sequencing. In this way, a lab can both guard against excessive (or inadequate) mutational bias that may skew results and also build up an empirical database of results that will guide future mutagenesis experiments.

ACKNOWLEDGMENTS

The research this document is based on was supported in part by the Countermeasures to Biological and Chemical Threats Program at the Institute for Advanced Technology of the University of Texas at Austin under contract number DAAD17-01-D-0001 from the U.S. Army Research Laboratory. In addition this work was supported by a grant from the Multidisciplinary Research Program of the University Research Initiative, No. DAAD19-99-1-0207, and also by grant number 1R01GM61789-01A1 from the National Institutes of Health. The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies of the U.S. government or the NIH unless so designated by other authorized documents.

REFERENCES

Index

A

Abasic sites, 11
ABI Prism genetic analyzer, 115
Acrylamide sequencing gel, 170
Acyclic amide enhancers, 59
Additive
potency of, 54
specificity, 54
Adenosine 5’ phosphosulfate (APS), 187
ADP, see Arbitrary degenerate primers
AFLP method, radioactive, 156
Agarose electrophoresis, 243
AIDS dementia, 212
Alkylation reaction, 126
Allele-specific hybridization, 98
Amplification
bias, 354
efficiency, 243
facilitators, 43
fragment libraries, 379
hemi-nested, 250
products, electrophoresis of, 52
Refractory Mutation System (ARMS), 99
single-molecule, 252
whole genome, 251, 351
Ancient DNA, extraction and amplification of, 5–9
PCR amplification, 7
phenol chloroform extraction protocol, 6
sample preparation, 6
silica and guanidinium thiocyanate extraction protocol, 6–7
Antibiotic resistance, 156
Antibody staining, 23
Antidigoxigenin alkaline phosphatase conjugate, 206
Antisense primer extension, 129
APEX, see Arrayed primer extension, primer design for large-scale multiplex PCR and APGAR scores, 212
APS, see Adenosine 5’ phosphosulfate
Arbitrary degenerate primers (ADP), 316
ARMS, see Amplification Refractory Mutation System
Arrayed primer extension (APEX), primer design for large-scale multiplex PCR and, 131–140
bioinformatics support for large-scale genomic PCR, 132–134
genome test, 132–134
multiplexing PCR primers, 134
optimal parameters for PCR primer design, 132
discussion of problems and practical hints, 138–139
APEX genotyping, 139
genome test, 138
primer design, 138
future directions, 139
PCR protocol, 135
SNP genotyping using APEX, 135–137
Arrays
discontinuous, 103
universal, types of, 102
Assay(s), see also GOOD assay
bar-coded, 102
discontinuous, 103
endpoint, advantage of, 225
error, 240
genotyping, 99
primer extension, 118
quality, relationship between primer specificity and, 133
quantitative PCR, 227
RT-PCR, real-time, 230
SNaPshot, 114–115, 118, 119
Ataxia with isolated Vitamin E Deficiency (AVED), 81
Autosticky PCR cloning, 87
AVED, see Ataxia with isolated Vitamin E Deficiency

B

Bacterial typing, 187, 193
Bar-coding, 102, 106
BIA, see Biospecific interaction analysis
BigDye octamer sequencing, 202
Bile salts, 41
Bilirubin, 41
Bioinformatics, 264
Biomphalaria glabrata, 279
Biopsy, 230, 259
Biorobotics technology, 243
Biosensor technology, see Surface plasmon resonance based biosensor technology for real-time detection of PCR products
Biospecific interaction analysis (BIA), 141, 143, 144, 149, 151
Biotin–streptavidin interactions, 145
Bisulfite conversion method, quantitative DNA-methylation analysis by, 175–185
bisulfite treatment and amplification, 177–180
drawbacks of method, 180
principle of bisulfite conversion reaction, 177
protocol for treatment of intact cells, 177–178
protocol for treatment of isolated DNA, 178
selective amplification of converted DNA, 178–179
troubleshooting, 179–180
materials, 176
bisulfite treatment, 176
purification of bisulfite-treated PCR product, 176
SNaPshot reaction and HPLC analysis, 176

381

www.taq.ir
quantiﬁcation, 180–184
analysis by SIRPH, 181–184
cloning and sequencing, 181

Bordetella pertussis, 38
Bovine serum albumin (BSA), 7, 8, 24, 43
Breast cancer cell line, estrogen-dependent, 273
BSA, see Bovine serum albumin
Buoyant density centrifugation, 39, 40

C

Campylobacter jejuni, 39
Cancers, gene expression in human, 261
Capillary Electrophoresis (CE), 98, 111
Capillary sequencers, 86
6-Carboxy-ßuorescine (FAM), 79
6-Carboxy-X-rhodamine (ROX), 79
CCD, see Charge-coupled device
cDNA
bacterial library of SSH PCR, 306
bands, recovery of, 293
digested, 300
double-stranded, 301
expression patterns, proﬁling of, 287
fly adapter-ligated, 295
ligation of digested, 292
microarray hybridization, 267
randomly ampliﬁed ﬁbroblast, 302
restriction enzyme digestion of, 292
synthesis, 291, 340
tester, 300
tiling arrays, 164, 172
cDNA, in situ ampliﬁcation of, 205–213
discussion, 210–212
budget, expertise, and effort required, 210
common pitfalls and troubleshooting tips, 210
other useful applications of method, 212
equipment and reagents, 206
instrumentation, 206
reagents, 206
experimental protocols, 207–210
in situ RT-PCR, 209
interpretation of data, 209–210
startup, 207
CE, see Capillary Electrophoresis
Cell(s)
counting, 23
fetal, 361
intact, protocol for treatment of, 177
isolation, 353
nuclei, counting of by ﬂow cytometry, 25
sorting, ﬂow cytometry, 27
visualization of captured, 34
Cellular remains, DNA extracted from, 5
CF, see Cystic ﬁbrosis
CGH, see Comparative genomic hybridization
Charge-coupled device (CCD), 112, 188
Chromatin, condensed, 220
Chromosome spreads, 217

PCR Technology: Current Innovations, Second Edition

Chromosome walking by inverse PCR, 323–328
discussion, 326–327
materials and methods, 325–326
analysis of PCR products, 326
circularizing DNA templates, 325
DNA templates, 325
inverse PCR, 325–326
primers, 325
results, 326
Cloning, see also DNA fragment, PCR-based techniques for cloning unknown; Virus sequences, cloning of unknown, by DNase treatment and sequence-independent single primer ampliﬁcation
autosticky PCR, 87
directional Topo, 92
enzyme-free, 87, 92
heterostagger, 87
ligation-independent, 92
restriction site-free, 87
Cloning method, universal restriction site-free, 87–92
discussion, 92
materials and methods, 88–91
cloning protocol, 88–91
enzymes and reagents, 88
Clostridium botulinum, 39
CNS dysfunction, 212
Collagen, 41
Comparative genomic hybridization (CGH), 354, 356, 357
Conﬁrmatory sequencing, 192
Cryostat microtome, 30, 31
Cyclic compounds, PCR enhancement and, 59
Cystic ﬁbrosis (CF), 142, 148, 150

D

Damaged DNA, long PCR from, 11–2
use of exonuclease III to enhance long PCR ampliﬁcation of damaged DNA templates, 15–19
common pitfalls, troubleshooting tips, and speciﬁc remarks, 18–19
mtDNA templates, 16
nuclear DNA templates, 16–18
use of long PCR to detect mtDNA damage, 12–15
DNA isolation, materials, and PCR instrumentation, 12–13
PCR protocol, 13–15
possible pitfalls, troubleshooting tips, and other remarks, 15
DASH, see Dynamic Allele Speciﬁc Hybridization
Database(s)
DNA, 117
ENSEMBL, 133, 138
gene expression, 261
genome, 155
DD, see Differential display
ddNTPs, see Dideoxynucleotides
Degenerate oligonucleotide primer (DOP), 104

www.taq.ir
Degenerate oligonucleotide primed PCR (DOP-PCR), 350, 352
Degenerate primers, 67
Denaturing High Performance Liquid Chromatography (DHPLC), 98
Dengue virus, 282, 283
De novo sequencing, 192
Deoxynucleotide triphosphates (dNTP), 88, 187, 202
Deoxyuridine triphosphates (dUTPs), 314
DEPC, see Diethyl pyrocarbonate
Designer FAFLP, 160
detection methods, wide-field, 106
DGE, see Differential gene expression
DHPLC, see Denaturing High Performance Liquid Chromatography
Dideoxynucleotides (ddNTPs), 114
chain-terminating, 197
fluorescently labeled, 137
Diethyl pyrocarbonate (DEPC), 38
Differential display (DD), 277, see also Infectious diseases, applications of differential display in Differential Display PCR, 299
Differential gene expression (DGE), 259
DiGeorge/velocardiofacial syndrome, 215
Dimethyl sulfoxide (DMSO), 13, 43, 51, 231, 327
diol(s)
low molecular-weight, 59
potency, 61
Directional Topo cloning, 92
Disease-associated genes, 193
Ditags, formation of, 271
Dithiothreitol (DTT), 289
DMSO, see Dimethyl sulfoxide
DNA, see also Ancient DNA, extraction and amplification;
Damaged DNA, long PCR from accessibility, 217
automated sequencer, 158
-binding protein, 43, 189
breakage, 12
chip arrays, 101
cloning, 92, 319
colonies, 105
conservation and storage of, 12
converted, selective amplification of, 178
databases, 117
denaturation, 12, 54
digest genomic, 157
:DNA hybridization, 141
double-stranded, 191, 342
extraction, 16, 41
 cellular remains, 5
 protocols, 37
 staining for, 31
 fingerprinting, direct genome walking and, 332
 fluorescent sequencing, 199
 genomic, cytosine methylation in, 175
 haploid, 129
 human cytomegalovirus, 38
 isolation, 12, 178
 isotopic sequencing, 199
 labeled molecular weight marker, 291
Legionella, 38
lesions, 19
melting, 56, 60
analysis, 52
cosolvent effects on, 61
temperature depression, 60
-methylation analysis, quantitative, see Bisulfite conversion method, quantitative DNA-methylation analysis by microarrays, 260
mitochondrial, 8, 11
damage, 12, 14
human identity testing and, 112
long PCR amplification of mouse, 13
one genome equivalent of, 246
polymerase(s), 13, 20
buffer, 82
fidelity, 74
polymerization activity of, 37
slippage, 357
pooling, 106, 167
precipitation, 378
purification, 98
quality, 128
reference, 254
repair enzymes, 19
samples
 genomic, 113
 heat-damaged, 17
separation, startup costs of method based on, 160
sequence, 7, 199
bisulfite-converted, 179
 variation, renaturation studies and, 97
shuffling, in vitro recombination via, 372, 377
single-stranded, 191
strand breaks, 19
template(s), 143, 165
pyrosequencing, 189
Tetrahymena, 325
 variation, measurement of, 111
visualization of labeled, 218
DNA fragment, PCR-based techniques for cloning unknown, 313–322
arbitrary primed-PCR–based method, 316–319
interspersed repetitive sequence-PCR–based methods, 314–316
ligation anchored-PCR–based methods, 316
PCR after intramolecular circularization, 313–314
DNA and RNA, microdissection of cells for analysis of, by PCR, 29–36
discussion, 34–35
 budget, expertise, and effort required, 34
other useful applications of method, 34–35
potential future developments, 35
examples, 30–34
common pitfalls and troubleshooting tips, 34
experimental protocol, 31–33
instrumentation and suppliers, 30
interpretation of data and results, 33
reagents and suppliers, 30–31
dNTP, see Deoxynucleotide triphosphates
www.taq.ir
examples and results, 71–73
materials and methods, 68–71
primer design, 66–68
Extension reactions, products of, 119

E

EC, see Endothelial cells
EDTA, see Ethylene diaminetetra-acid
Electronic OST (eOST), 197
Electrophoresis
agarose, 243
Capillary, 98, 111
gel, 56, 79, 98, 248
high-resolution, 262
result, SNAPPshot primer extension assay, 119
ELISAs, 65
ELT PCR, see Expand Long Template PCR
Endothelial cells (EC), 24, 25
Endpoint assays, advantage of, 225
Enterovirus RNA, 38
Enzyme(s)
digestions, 277
DNA digestion of, 156
-free cloning, 87, 92
template-generating, 327
eOST, see Electronic OST
Escherichia coli
exonuclease III, 15
O157:H7, 39
peptidyl-tRNA hydrolase, 90
repair enzyme of, 12
XL1-Blue, 198
EtBr, see Ethidium bromide
Ethidium bromide (EtBr), 71
Ethylene diaminetetra-acid (EDTA), 6, 115, 176
Exonuclease III
beneficial effect of, 17
damaged nuclear DNA templates and, 16
enhancing effect of, 19
Escherichia coli, 15
primer degradation and, 18
Expand Long Template (ELT) PCR, 363
Experimental design, optimizing PCR with aid of, 65–75
optimization of reaction mixture and temperature cycle, 68–75
discussion, 73–75

F

FACS, see Fluorescence activated cell sorting
FACSort flow cytometer, 25
FAFLP, see Fluorescent amplified fragment length polymorphism genotyping
FAM, see 6-Carboxy-fluorescence
FCS, see Fluorescence Correlation Spectroscopy
Fetal cells, 361
Fibrinolytic system, regulator of, 23
Fibroblast cDNA, 302
FISH, see Fluorescent in situ hybridization
FITC, see Fluorescin isothiocyanate
Flow cytometry, see RT-PCR and flow cytometry, quantitative mRNA analysis in small cell samples by
Fluorescin isothiocyanate (FITC), 25
Fluorescence
activated cell sorting (FACS), 25, 379
-based real-time chemistries, 226
Correlation Spectroscopy (FCS), 105
ethidium bromide, 71
Polarization (FP), 101
resonance energy transfer (FRET), 52, 100
technology, multicolor, 111
tryptophan, 60
Fluorescent amplified fragment length polymorphism (FAFLP) genotyping, 155–161
background, 155
discussion, 160
FAFLP, 156
materials and methods, 157–158
equipment, 157
materials, 157
method, 157–158
results, 158–159
Fluorescent dyes, detection of, 118
Fluorescent fragment analysis, 160
Fluorescent in situ hybridization (FISH), 215, 357
Fluorescent labeling, 80, 84, 85
Fluorescent primers, universal, 81
Fluorescent tag, attachment of, 79
Fly adapters, 290, 292
Forensic identification, 192
Forensic samples, fingerprinting of, 245
FP, see Fluorescence Polarization
FRET, see Fluorescence resonance energy transfer
Fungal typing, 187, 193

G

GADA V , see Global Analysis of DNA Allelic Variations
GADPH, see Glyceraldehyde-3-phosphate dehydrogenase
Gel electrophoresis, 56, 79, 98, 248
Index

Genbank, 272

Gene(s)
  all or nothing differentially expressed, 301, 303, 305
differentially displayed, 282
enrichment of ratio differentially expressed, 304
GAPDH, 364, 366
identification of, 281
protein coding, 288
target discovery, 262

GeneCalling, 261, 262, 264
Gene expression and discovery, 259–264
applications to drug target discovery, 262–264
functional genomics, 262–264
toxicogenomics, 264
closed systems, 260–261
open systems, 261–262

Genetically modified organisms (GMO), 149
Genetic analysis, see High-throughput genetic analysis through multiplexed PCR and multicapillary electrophoreses
Genetic analyzer, ABI Prism, 115
Genetic mutations, diagnosing complex, 152

Genome
  databases, 155
  mapping, 255, 267
  project, Drosophila, 329, 332
  -sampling methods, 155
  template, 355
test, 132, 133, 138
walking, 329

Genomic mining, 263

Genotools* SNP manager software, 123
Genotyping, see also Fluorescent amplified fragment length polymorphism (FAFLP) genotyping; High-throughput genotyping, PCR and APEX, 139
assays, major platforms for performing, 99
software, 118
time, comparison of, 120

Genzyme Molecular Oncology, 268

Geobacter spp., 38

Global Analysis of DNA Allelic Variations (GADAV), 163–174
background and aims, 163–164
discussion, 171–172
  future developments and potential applications, 171–172
summary, 172
materials and methods, 164–169
experimental protocol, 167–169
materials, 164–167
outline of GADAV strategy, 164
results, 169–171
  enrichment of Mn DNA duplex from digested pBR322 Pm fragments, 169
scanning of polymorphic fragments in human cDNA pool, 169–171
Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), 229, 306, 362
Glycolipid transfer protein (GTP), 52, 56
GMO, see Genetically modified organisms

GOOD assay, 121–130
discussion, 128–129
  budget, expertise, and effort required, 128
  other useful applications, 128–129
method, 122–128
  common pitfalls and troubleshooting, 127–128
  examples, 125–127
instrumentation and suppliers, 122–123
interpretation of data, 127
reagents and suppliers, 124–125
software, 123

GTP, see Glycolipid transfer protein

H

Haplotype tag SNPs (htSNPs), 131
HAPPY mapping, 255, 256
Hardy–Weinberg test, 133
HBV, see Hepatitis B virus
HCV, see Hepatitis C virus
Helicobacter hepaticus, 40
Hemoglobin, 41
Hepatitis B virus (HBV), 316
Hepatitis C virus (HCV), 235
Heterohybrids, 168, 171
Heterostagger cloning, 87
HEX, see Hexachloro-6-carboxy-fluorescine
Hexachloro-6-carboxy-fluorescine (HEX), 79
HGP, see Human Genome Project
High-resolution printing, 102
High-throughput genetic analysis through multiplexed PCR and multicapillary electrophoreses, 111–120
materials and methods, 113–117
  ABI Prism® 310 genetic analyzer, 115
  ABI Prism® 3100 genetic analyzer, 116
data analysis, 117
genomic DNA samples, 113
PCR amplification and typing assays, 113–115
sample preparation for data collection, 115
results and discussion, 117–119
SNP typing, 118–119
STR typing, 117–118
High-throughput genotyping, PCR and, 97–110
future perspectives, 103–107
  alternatives to PCR for large-scale studies, 105–106
  concluding remarks, 106–107
  new ways for using PCR in large-scale studies, 103–105
platforms currently available, 98–103
array methods, 101–103
homogeneous methods, 99–101
separation methods, 98–99
HIV, see Human immunodeficiency virus
HM, see Hydraulic micromanipulation
Homohybrids, 171
Hotstart enzyme system, 249
Housekeeping genes, 26, 227
HPLC, 248
htSNPs, see Haplotype tag SNPs

www.taq.ir
Human β-globin gene, 17
Human cytomegalovirus DNA, 38
Human Genome Project (HGP), 264
Human immunodeficiency virus (HIV), 192, 212, 235
Human umbilical vein endothelial cells (HUVEC), 24, 27
Humic materials, 41, 42
HUVEC, see Human umbilical vein endothelial cells
Hybridization
cDNA microarray, 267
determination of optimal conditions for, 144
DNA:DNA, 141
in situ, 205, 215, 357
reaction, efficiency of, 98
subtractive, 299, 338
Hydraulic micromanipulation (HM), 30, 34
Hydrocarbons, reaction-inhibiting, 69
Hydroxyapatite, 8
ICD, see Identity-by-Descent
Identity-by-Descent (IBD), 164
IHJ, see Integrant-host junction
IL-2R, see Interleukin-2 receptor
Immunoglobulin G, 41
Immunochemistry, 209
Immunochemistry, 209
Immunomagnetic separation (IMS), 39
Immunoscreening, 338
IL-2R, see Interleukin-2 receptor
Interspersed repetitive sequence (IRS), 314
Infectious diseases, applications of differential display in,
277–285
applications in area of infectious diseases, 278–279
definition and theoretical aspects, 277–278
differentially displayed genes during dengue virus infection, 282–283
materials and methods, 280–282
differential display and isolation of DNA bands for
sequence analysis, 280–281
identification of genes, 281
preparation of total cellular RNA, 280
reverse transcription, 280
semiquantitative RT-PCR and real-time SYBR
Green® RT-PCR, 281–282
optimization of DD, 279–280
In situ hybridization, 205, 215, 357
Integrant-host junction (IHI), 313
Interleukin-2 receptor (IL-2R), 227
Interspersed repetitive sequence (IRS), 314
Inverse PCR (iPCR), 313, 332, see also Chromosome walking by inverse PCR
In vitro mutagenesis, 92
Ion mode detection
negative, 126
positive, 125
iPCR, see Inverse PCR
IRS, see Interspersed repetitive sequence
Magnetic capture hybridization (MCH), 39
MALDI, see Matrix-Assisted Laser Desorption Ionization
Mapping, HAPPY, 255, 256
Massively Parallel Signature Sequencing (MPSS), 103
Mass Spectrometry (MS), 98, 100, 122, 127
Matrix-Assisted Laser Desorption Ionization (MALDI), 98
mass spectrometry, 122, 127, 128, see also GOOD assay
spectra, automatic recording of, 123
MCH, see Magnetic capture hybridization
Metallothionein, 164
Methylation, imprinted patterns of, 184
Methylation, 181
2-Methyl-2,4-pentanediol (MPD), 56
MicroBeta Liquid Scintillator Counter, 26
Microchip Affymetrix analysis, 278
Microdissection, features of different methods for, 33
Microsatellite analysis, 82, 83
Index

labeling, 80, 82
markers, 366
-specific markers, primers for, 81
Micro Total Analytical Systems, 103
Mitochondrial DNA (mtDNA), 8, 11, 112
damage, 12, 14
long PCR amplification of mouse, 13
templates, damaged, 16
Molecular beacons, 100
Molecular biology, major task of modern, 163
MPD, see 2-Methyl-2,4-pentanediol
MPSS, see Massively Parallel Signature Sequencing
mRNA, meaningful quantification of, using real-time PCR, 225–233
critical factors, 228–231
normalization, 228–229
quantification and quality assessment of RNA, 229–230
relevance, 230–231
reproducibility, 230
reverse transcription, 231
types of quantitative PCR assays, 227–228
absolute quantification, 227–228
relative quantification, 227
MS, see Mass Spectrometry
mtDNA, see Mitochondrial DNA
MTV, see Multistep-touchdown vectorette
Multicapillary electrophoresis, see High-throughput genetic analysis through multiplexed PCR and multicapillary electrophoreses
Multicolor fluorescence technology, 111
Multiplex microsatellite labeling, 82
Multiplex PCR, 112
Multiplex STR typing kits, commercially available, 113
Multistep-touchdown vectorette (MTV), 316
Mutagenesis and recombination, in vitro evolution through PCR-mediated, 371–380
in vitro recombination via DNA shuffling, 372–376
materials and methods, 376–379
in vitro recombination via DNA shuffling, 377–379
mutagenic PCR, 376–377
mutagenic PCR, 372
results and discussion, 379
Mutation, random, 371
Mycobacteria tuberculosis, 278
Mycoplasma iowae, 40
Myoglobin, 41

N

Neisseria gonorrhoeae, 158, 159
Nonhomologous random recombination (NRR), 375
NRR, see Nonhomologous random recombination
Nucleic acid extraction, 29
Nucleic acids, absolute quantification of, by RT-PCR, 235–244
competitive RT-PCR, 238–239
data analysis with PCRFIT program, 239–242
assessment of assay error, 240–241
determination of copy number, 242
input data, 239
output data, 239–240
discussion, 242–243
experimental problem, 236
generation of RNA mimic and reference target, 237–238
in vitro transcription, 238
ligation-PCR, 238
primer design, 237
RT-PCR, 237
Nucleotide(s)
mutagenesis via incorporation of modified, 376
variations, global scanning of, 165, 172
wrong natural, 372

O

OCT, see Optimal cutting temperature
Octamer-primed cycle sequencing, 197–203
discussion, 203
materials and methods, 198–200
analysis of DNA sequence read-length and data accuracy, 200
octamer-primed DNA sequencing, 199–200
preparation of sequencing extension products for analysis, 200
primer selection and synthesis, 198
purification of plasmid DNA templates for octamer-primed sequencing, 198–199
results, 200–203
development of octamer-primer library, 200–201
optimization of sequencing reaction chemistry, 202–203
primer identification software, 201–202
Octamer sequencing technology (OST), 197
ODN, see Oligodeoxyribonucleotide
Oligodeoxyribonucleotide (ODN), 142
Oligonucleotide(s), 289
adaptors, 172
allele-specific cleavage of, 122
amino-modified, 124
APEX and, 136
bar-code arrays, 102
desalted, 68
sequences of, used as primers, 157
synthetic, 339
Optimal cutting temperature (OCT), 31
Organic solvents, PCR enhancement by, 54
OST, see Octamer sequencing technology

P

PAI-1, see Plasminogen activator inhibitor-1
Partial inverse PCR (PI-PCR), 314
PBS, see Phosphate-buffered saline
PCR (polymerase chain reaction), 23, 111, 225, 235, 349
alternatives to, 105
amplification, 171, 292
ancient DNA, 7

www.taq.ir
effects of cosolvents on, 53
failure of, 37, 180
modeling of, 60
priming site for, 269
spermidine and, 12
assays, quantitative, 227
blanks, 5
carryover, 248
cloning
autosticky, 87
strategies, most common, 67
conditions
optimizing, 179
pyrosequencing, 188
contamination, 247, 290
degenerate oligonucleotide primed, 350, 352
Differential Display, 299
Expand Long Template, 363
facilitators, 40
hotstart, 193
immobilized, 104
inhibitors, 15, 43, 229
in situ, 206
inverse, 323
large-scale multiplex, see Arrayed primer extension,
prime design for large-scale multiplex PCR
and ligation-, 238
multiplex, 112, 250
mutagenic, 372, 376
optimization
buffers, 378
experimental design for, 65
spreadsheets, 69
partial inverse, 314
PEP-, 363
prime(s)
design, optimal parameters for, 132
multiplexing, 134
pyrosequencing, 190
product(s), 151, see also Surface plasmon resonance
based biosensor technology for real-time
detection of PCR products
analysis of, 14, 17
bisulfite-treated, 176, 181
direct sequencing of, 296
fluorescent labeling of, 80
immobilized, 147
rare-earth metal ion and, 87
real-time detection of, 146
SPR-based BIA of, 151
protocol
fluorescent labeling multiplex, 84
four-primer long, 13
setting up of, 15
RACE, 325
real-time quantitative, 260, 263, 299
RT-, see RT-PCR; RT-PCR and flow cytometry,
quantitative mRNA analysis in small cell
samples by
sample preparation methods, 38
single-cell, 33, 361
Suppression, 165, 301
tagged, 350, 352
thermal asymmetric interlaced, 316, 318, 321
tube, 215, 236
two-phase, 248
PCR-enhancing compounds, novel, 51–63
discussion, 58–62
identification of outstanding PCR enhancers, 58–60
mechanisms of PCR enhancement, 60–62
materials and methods, 52–53
DNA melting analysis, 52–53
PCR reactions, 52
Taq polymerase activity and thermostability
analysis, 53
results, 53–58
DNA melting, 56–57
PCR enhancement by organic solvents, 54–56
Taq polymerase thermostability and activity, 57–58
templates used and definition of terms, 53–54
PECAM, see Platelet/endothelial cell adhesion molecule
PEG, see Polyethylene glycol
PEN, see Polyethylene naphthalene
PEP, see Primer extension preamplification
Peptide nucleic acid (PNA), 142, 143, 149
Personalized medicine, 131
PGD, see Preimplantation genetic diagnosis
Phenol chloroform extraction, 6
Phosphate-buffered saline (PBS), 176
Phospholmager, 169
Photolithography, 102
PHRED, 198, 201
Pico LabChip®, 229
PI-PCR, see Partial inverse PCR
Plasminogen activator inhibitor-1 (PAI-1), 23
Plasmodium falciparum, 279
Platelet/endothelial cell adhesion molecule (PECAM), 126
PNA, see Peptide nucleic acid
Poisson distribution, 246
Polyacrylamide gels, denaturing, 197
Polyethylene glycol (PEG), 40, 42
Polyethylene naphthalene (PEN), 30
Polymerase
chain reaction, see PCR
extension, 51
thermostability, 57, 59, 60
Polyphenols, 42
Polysaccharides, 41
PPi, see Inorganic pyrophosphate
PraderWilli/Angelman syndrome, 215
Pregnancy, termination of, 367
Preimplantation genetic diagnosis (PGD), 349
Prenatal diagnosis, 361
Pre-PCR processing strategies, 37–47
DNA polymerases, 40
equivalents of, 41–42
PCR facilitators, 40–43
sample preparation, 38–40
biochemical methods, 38–39
enrichment methods, 39
immunological methods, 39

www.taq.ir
Index

physical methods, 39–40
Primed in situ labeling (PRINS), 215
Primer(s)
  arbitrary degenerate, 316
  bad, 134
  concentration, 70
  degenerate, 67, 104
  design, 66, 138
  dimers, 66
  extension
    antisense, 129
    assay, 118
    reaction, 126, 127
  good, 134
  identification software, 201
  melting temperatures, 66
  microsatellite-specific markers, 81
  mismatched, 66
  PCR, pyrosequencing, 190
  –primer interactions, 139
  quality, 127
  sequences, 157, 237
  SNP extension, 114
  SNuPE, 183
  specificity, relationship between assay quality and, 133
  universal, 80, 81, 86, 210
Primer extension preamplification (PEP), 251, 350, 353, 361–368
  discussion, 367
  example, 362–366
  analysis of PEP efficiency using real-time quantitative PCR, 363–364
  interpretation of results, 364–366
  micromanipulation of single cells using LMM/LPC, 362–363
  PEP-PCR protocol, 363
  pitfalls and troubleshooting tips, 366
  preparation slides, 362
PRIMER 3 software, 123
PRINS, see Primed in situ labeling
PRINS, detection of single-copy sequence on human metaphase by cyclic, 215–221
  discussion, 218–220
  experimental protocol, 217–218
  instrumentation, 216
  reagents, 216
Prostate-specific membrane antigen (PSM), 52
Protease digestion
  determining optimal, 208
  inadequate, 210
Protein
  coding genes, 288
  expression, gene cloning for, 68
  –nucleic acid crosslinks, 207
  ribosomal, 170
Proteinase K digestion, 5, 68
PSM, see Prostate-specific membrane antigen
Pyrogram, 190
Pyrosequencing technology, 187–195
  discussion, 193
materials and methods, 188–190
  PCR conditions, 188–189
  preparation of DNA template, 189–190
  pyrosequencing, 190
  results, 190–193
  applications, 192–193
  choice of PCR primers, 190–191
  DNA preparation, 191
  pyrosequencing, 191–192

Q

Qbiogene, 18, 19, 20

R

RACE PCR, 325
Radioisotopic sequencing, 201
Randomly amplified polymorphic DNA (RAPD), 40
Random-priming recombination (RPR), 373
RAPD, see Randomly amplified polymorphic DNA
Rare-earth metal ion
  incubation of PCR products with, 87
  treatment, role of, 89
RCA, see Rolling Circle Amplification
RCF, see Relative centrifugal force
RDA, see Representative differential analysis
Reaction dropout, 71
READS, see Restriction endonucleolytic analysis of differentially expressed sequences
Real-time quantitative polymerase chain reaction (RTQ-PCR), 260, 263, 299
Recombination, see Mutagenesis and recombination, in vitro evolution through PCR-mediated
Reference template, 252, 253
Relative centrifugal force (RCF), 7
Representative differential analysis (RDA), 164, 301, 338
Restriction
  digestion, 89, 341
  enzymes, fly adapters for, 295
  fragment length polymorphism (RFLP), 40, 97, 155
  site-free cloning, 87
Restriction endonucleolytic analysis of differentially expressed sequences (READS), 261, 287–298
  discussion, 296–297
  experimental protocol, 290–296
  autoradiography, 293
  cDNA synthesis, 291–292
  data documentation, 294
  general precautions, 290
  ligation of digested cDNA with fly adapter, 292
  PCR amplification, 292–293
  processing of PCR-amplified samples for sequence analysis, 294
  reagents to be prepared in advance, 290–291
  recovery of cDNA bands from gel and PCR amplification, 293–294
  restriction enzyme digestion of cDNA, 292
  sequencing gel analysis, 293
technical comments, 295–296
materials and methods, 288–290
enzymes, 289
oligonucleotides, 289–290
reagents, 288–289
Reverse transcriptase (RT), 23, 205, 225, 235, 279, 280
RFLP, see Restriction fragment length polymorphism
Ribosomal RNA (rRNA), 229
RNA, see also DNA and RNA, microdissection of cells for analysis of by PCR
enterovirus, 38
ex vivo, 41, 341
staining for, 31
fingerprinting, 299
isolation methods, 26
mimic, 237
preparation of total cellular, 280
quality assessment of, 229
ribosomal, 229
sample, READS analysis of, 297
viruses, 211, 340
Rolling Circle Amplification (RCA), 105
ROX, see 6-Carboxy-X-rhodamine
RPR, see Random-priming recombination
rRNA, see Ribosomal RNA
RT, see Reverse transcriptase
RT-PCR, 23, 205, see also Nucleic acids, absolute quantification of, by RT-PCR; RT-PCR and flow cytometry, quantitative mRNA analysis in small cell samples by analysis, estimating precision of, 26 assays, real-time, 230 competitive, 238, 241 in situ, 207, 209, 210 preparation of cell lysates for, 25 quantitative, 277 semiquantitative, 282
RT-PCR and flow cytometry, quantitative mRNA analysis in small cell samples by, 23–28 materials and methods, 24–26 counting of cell nuclei by flow cytometry, 25 endothelial cell sorting by flow cytometry, 24–25 isolation and culture of endothelial cells, 24 preparation of cell lysates for RT-PCR, 25 RT-PCR, primers, and internal standard, 26 results and discussion, 26–27
RTQ-PCR, see Real-time quantitative polymerase chain reaction
SAGE, see Serial analysis of gene expression, detection of differential gene expression changes using
Salmonella Mbandaka, 39
SAP, see Shrimp alkaline phosphatase
Scanning tunneling microscope (STM), 106
Schistosoma sp., 279
SDS, see Sodium dodecyl sulfate
SelfSeal reagent, 206
Semiconductor quantum dot nanoparticles, 103
Sensiscrypt® RT, 279, 280
Sensorgram, 147, 150
Sequence-tagged sites (STTs), 256
Sequencing
BigDye octamer, 202
certification, 192
de novo, 192
genetic analysis, 293
information, collecting of, 119
radioisotopic, 201
reaction chemistry, 202
Serial analysis of gene expression (SAGE), detection of differential gene expression changes using, 267–275
discussion, 273–274
applications of SAGE, 273
development of SAGE technology, 273
using SAGE to annotate human genome, 273–274
materials and methods, 268–273
interpretation of data, 272–273
preparation of SAGE libraries, 268–272
reagents, 268
SERS, see Surface Enhanced Raman Scattering
Short tandem repeat (STR), 8, 40, 112, 121
analysis, steps in sample processing for, 114
genotyping result, 118
typing, 112, 117–118
Shrimp alkaline phosphatase (SAP), 122, 124, 135, 181
Signal-to-noise level (SNL), 70, 71
Signal-to-noise ratio, 123, 172, 242
Single-fluorophore detection, 106
Single-molecule PCR, 245–257
applications and theory, 252–256
determination of copy numbers, 254–255
mapping, 255–256
reference reagents, validation and quantitation of single-molecule amplification, 252–254
basic tool kit, 246–251
basic single-molecule amplification by nested two-phase PCR, 248–250
contamination control, 247
expressing amounts and concentrations of DNA, 246
hemi-nested amplification, 250
highly multiplex single-molecule PCR, 250
PEP and other nonspecific amplification methods, 250–251
preparation and handling of template DNA, 247
simple statistics, 246–247
discussion, 256
Single nucleotide polymorphisms (SNPs), 97, 112, 121, 187
analysis, steps in sample processing for, 114
extension primers, 114
genotyping production, 128
typing, SNaPshot assay, 114, 118, 119
verification, 127
Single-stranded conformation polymorphism (SSCP) analysis, economic fluorescent labeling of PCR products for microsatellite- and, 79–86
discussion, 85–86

www.taq.ir
Index

materials and methods, 81–83
DNA templates, 81
multiplex microsatellite labeling, 82–83
SCCP analysis, 83
results, 83–85
microsatellite analysis, 83–84
SCCP analysis, 84–85
Single-stranded DNA (ssDNA), 191
SIRPH, see SNuPE protocol using ion pair reverse phase high performance liquid chromatography
Smart Cycler®, 38
SNaPshot assay, 114, 118, 119
SNL, see Signal-to-noise level
SNPs, see Single nucleotide polymorphisms
SNuPE protocol using ion pair reverse phase high performance liquid chromatography (SIRPH), 176, 180, 181
advantages and drawbacks, 184
analysis of results, 183
guidelines for designing SNuPE primers, 183
protocol outline, 181–183
troubleshooting, 183–184
Sodium dodecyl sulfate (SDS), 176
Software
BLAST, 133, 138, 237, 281
Genotools® SNP manager, 123
genotyping, 118
MEGABLAST, 133, 138
mfold, 146
PCRFIT, 236, 239
PHRED, 198, 201
PRIMER 3, 123
primer identification, 201
SSAHA, 133, 138
WaveMaker, 181, 183
Southern Blot Analysis, 171
SPEX Fluoromax-3 fluorometer, 53
SPR, see Surface Plasmon Resonance
SSCP analysis, see Single-stranded conformation polymorphism (SSCP) analysis, economic fluorescent labeling of PCR products for microsatellite and ssDNA, see Single-stranded DNA
SSH PCR, see Suppression Subtractive Hybridization PCR
Staggered Extension Process (StEP), 373
Staphylococcus aureus, 40
Stem cell factor-dependent chemotaxis, 35
StEP, see Staggered Extension Process
STM, see Scanning tunneling microscope
STR, see Short tandem repeat
STSs, see Sequence-tagged sites
Stutter artifact, 357
Subtractive hybridization, 299
Superscript II Reverse Transcriptase, 289
Suppression Subtractive Hybridization PCR (SSH PCR), 299–307
discussion, 304–306
theoretical considerations, 304–306
troubleshooting, 306
materials and methods, 300–301
first hybridization, 300
ligation of adapters to tester cDNA, 300
PCR amplification, 301
preparation of cDNA, 300
primers, 300
second hybridization, 301
results, 301–304
efficacy of SSH PCR, 301–304
outline of SSH PCR, 301
Surface Enhanced Raman Scattering (SERS), 105
Surface Plasmon Resonance (SPR), 141
Surface plasmon resonance based biosensor technology for real-time detection of PCR products, 141–153
discussion, 149–152
advantages of SPR-based BIA for characterization of PCR products, 149
future perspectives, 152
problems and possible solutions, 151–152
materials and methods, 143–145
DNA templates and PCR conditions, 143
primers, molecular probes, and target oligonucleotides, 143
SPR-based BIA, 143–145
results, 146–149
characterization of PCR products immobilized on flow cells, 147–149
choice of primers and probes, 146
injection of asymmetric PCR products to flow cells carrying specific probes, 149
use of peptide nucleic acids as molecular probes, 149
Surgical pathology, 209

T
Tagged-PCR (T-PCR), 350, 352
Taguchi method, 71
critical factors in, 73
optimization, 72
optimum parameters obtained with, 74
precautions for PCR optimization with, 70
reactions optimized by, 68
signal-to-noise levels, 70
spreadsheet program, 69
TAIL-PCR, see Thermal asymmetric interlaced-PCR
Taq DNA polymerase, see Thermus aquaticus DNA polymerase
TDGs, see Thymine DNA glycosylases
Template switching, 357
Test tube evolution, 371
Tetrahymena, 325
Thalassemia mutation, 149
Thermal asymmetric interlaced (TAIL)-PCR, 316, 318, 321, 329
Thermal decay, 60
Thermus aquaticus (Taq) DNA polymerase, 372
activity, 53
thermostability, 57, 59
Thymine DNA glycosylases (TDGs), 164, 166
Tissue biopsy, 259
TOGA, see Total gene expression analysis

www.taq.ir
TOPO TA Cloning vector, 169
Total gene expression analysis (TOGA), 261
Toxicogenomics, 264
T-PCR, see Tagged-PCR
Tryptophan β-fluorescence, 60
Tube PCR, 215
Tumors, classification of human, 287
Tyramide Signal Amplification System, 216

UDG, see Uracil DNA glycosylase
UEA I, see Ulex europaeus I agglutinin
UFW, see Universal Fast Walking
Ulex europaeus I agglutinin (UEA I), 24
UNG, see Uracil-N-glycosylase
Universal arrays, types of, 102
Universal energy-transfer labeled, 210
Universal Fast Walking (UFW), 329
Universal primer, 80, 81, 86
Universal restriction site-free cloning, 87
Uracil DNA glycosylase (UDG), 314
Uracil-N-glycosylase (UNG), 137
Urobilogens, 41

Virus(es)
association between disease and newly discovered, 344
dengue, 282, 283
extraction, 42
hepatitis B, 316
hepatitis C, 235
human immunodeficiency, 235
isolation of, 337
RNA, 211
sequences, 343
typing, 187, 193
Virus sequences, cloning of unknown, by DNase treatment and sequence-independent single primer amplification, 337–345
discussion, 344
discussion, 344
discussion, 344
example, 339–344
experimental protocol, 340–343
interpretation of data, 343–344
key reagents, 339–340

W
WaveMaker software, 181, 183
West Nile virus, 38
WGA, see Whole genome amplification
Whole genome amplification (WGA), 251, 349–359
discussion, 356–358
materials and methods, 352–354
assessment and analysis of WGA products, 353–354
chromosomal analysis using comparative genomic hybridization, 354
primer extension preamplification, 353
sample preparation and cell lysis, 352
tagged PCR, 352–353
WGA using ALU-PCR, 353
WGA using degenerate oligonucleotide primed PCR, 352
results, 354–356
coverage of template genome and overall yields, 355–356
cytogenetic data from WGA products, 356
fidelity of WGA, 356
fragment size and amplification bias, 354–355
Wide-field detection methods, 106

X
X-ray film scanner, 294

Y
Yersinia enterocolitica, 39, 40

Z
Zip-codes, 102
PCR TECHNOLOGY
Current Innovations

Featuring breakthroughs in PCR technology and applications developed since the publication of the bestselling first edition, *PCR Technology: Current Innovations, Second Edition* provides the essential tools for developing innovative approaches to using this leading technology of the genomics revolution. The book includes theoretical considerations, discussions, and a selection of state-of-the-art techniques for mutation studies, clinical diagnosis, and the detection of food-borne pathogens. With detailed presentations of PCR protocols, this new edition discusses the preparation of PCR experiments, includes examples of analytical PCR divided into qualitative and quantitative applications, and explores preparative methods that address DNA generation for further analysis and *in vitro* evolution.

**Features**

- Provides up-to-date, essential information on PCR technology and applications
- Covers a selection of widely used procedures and approaches for qualitative as well as quantitative queries
- Explains parameters governing PCR reaction outcomes
- Presents easy-to-follow, step-by-step protocols
- Includes primer design, enzyme choices, precautions to avoid false positives, and more

Covering the current range of PCR strategies and uses, this volume is a fundamental reference for scientists new to PCR technology and a source of up-to-date applications for those familiar with the basic methods. It contains critical information for practitioners in a wide variety of fields, including forensics, molecular biology, clinical DNA diagnostics, botany, and paleontology.

**CRC PRESS**

www.crcpress.com