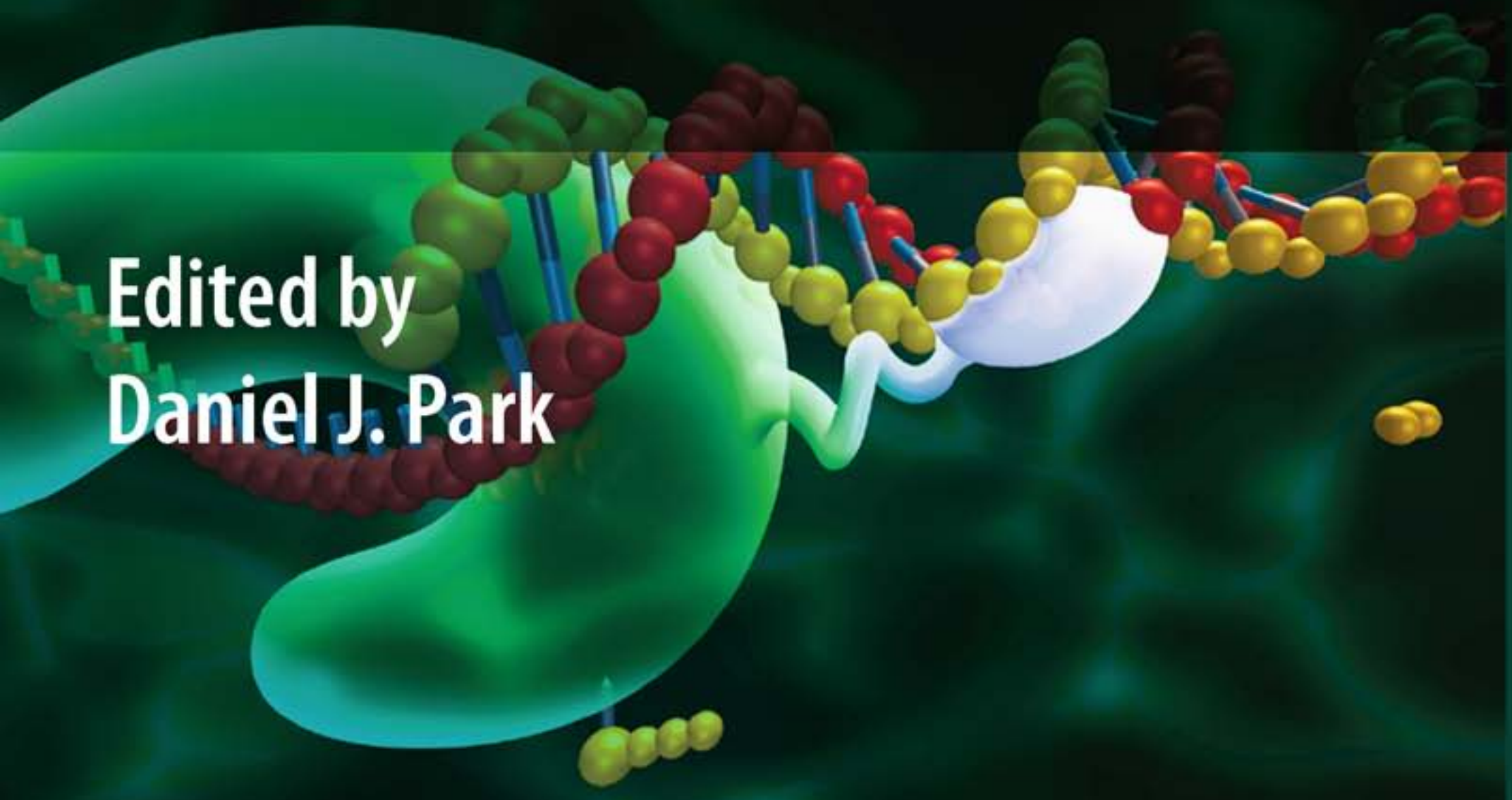
Springer Protocols

Methods in Molecular Biology 687

PCR Protocols

Third Edition

Edited by
Daniel J. Park



Humana Press

METHODS IN MOLECULAR BIOLOGY™

Series Editor
John M. Walker
School of Life Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For other titles published in this series, go to
www.springer.com/series/7651

PCR Protocols

Third Edition

Edited by

Daniel J. Park

Genetic Epidemiology Laboratory, Department of Pathology,
The University of Melbourne, Parkville, VIC, Australia

Humana Press

Editor
Daniel J. Park, Ph.D. (CANTAB)
Genetic Epidemiology Laboratory
Department of Pathology
The University of Melbourne
Parkville, VIC
Australia
djp@unimelb.edu.au

ISSN 1064-3745 e-ISSN 1940-6029
ISBN 978-1-60761-943-7 e-ISBN 978-1-60761-944-4
DOI 10.1007/978-1-60761-944-4
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: PCN applied for

© Springer Science+Business Media, LLC 2011
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Humana Press, c/o Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.
While the advice and information in this book are believed to be true and accurate at the date of going to press, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Cover illustration: Schematic structure of Phusion® High-Fidelity DNA Polymerase, kindly provided by Finnzymes Oy, a Thermo Fisher Scientific Company.

Printed on acid-free paper

Humana Press is part of Springer Science+Business Media (www.springer.com)

Preface

Since the concept of PCR was first described, PCR has taken on a central role in many areas of biological research and diagnostics. This continues today, including its application to the latest generation genome-wide analysis and high-throughput sequencing platforms. Although other methods have been described to enable the exponential amplification of nucleic acids, such as NASBA, LAMP, and LCR, PCR remains unparalleled in terms of its flexibility and robustness.

From the basic concept of using a pair of primers and polymerase to amplify a defined region of nucleic acid, a multitude of derivative techniques and applications have emerged. Numerous developments have occurred to improve on early iterations of PCR, including the identification of thermostable DNA polymerases which exhibit a range of properties to suit given applications. The use of such enzymes offer tailored and improved processivity, fidelity, and specificity. Combined with variations in primer design, pre-processing techniques, thermocycling settings, use of substrates and substrate analogues, and methods of amplicon detection, to name but a few parameters, PCR offers an extremely powerful and diverse portfolio of methodologies. PCR can be nested or not, stepped up or stepped down, one or two-sided, stringent or "relaxed," long or short, solid or aqueous, single or multiplex, blunt or sticky, and error prone or high-fidelity. PCR can be suppressed, looped or inverse, can shuffle, use adapters, fluoresce, can be read in real time, and can be used to detect a variety of analytes, including RNA, DNA, or protein.

Previous publications have introduced PCR along with a range of established protocols. Given the almost unlimited potential of "PCR protocols" and with the aim of somewhat limiting the retreading of old ground, this volume selects relatively recently described tools and tricks contributed by field-leading authors for the value that they are considered to add to more generally established methods. This is attempted along with the spanning of a range of core applications: PCR cloning and sequencing, expression, copy number or methylation profile analysis, "DNA fingerprinting", diagnostics, protein engineering, and interaction screening. A chapter highlighting workflow considerations and contamination control is intended as a "before you begin" item to stress the importance of addressing these issues before proceeding to follow any PCR protocol.

While *PCR Protocols, Third Edition* is collated in the form of a series of independent methods, core principles presented in any given chapter have broader application. Read in this light, along with alternative approaches to consider listed in the introductory sections of chapters and the special considerations offered in the Notes sections, it is hoped that this volume offers more than the sum of its parts.

Parkville, VIC, Australia — *Daniel J. Park*

Contents

Preface . . . . . *v*
Contributors . . . . . *ix*

PART I PCR HUSBANDRY

1 Setup of a PCR Laboratory 3
Zaheer Khan

PART II CLONING AND SEQUENCING

2 Long-Range PCR with a DNA Polymerase Fusion 17
Holly H. Hogrefe and Michael C. Borns

3 Isolation of Genomic Insertion Sites of Proviruses Using Splinkerette-PCR-Based Procedures 25
Bin Yin

4 Lariat-Dependent Nested PCR for Flanking Sequence Determination 43
Daniel J. Park

5 CODEHOP PCR and CODEHOP PCR Primer Design 57
Jeannette P. Staheli, Richard Boyce, Dina Kovarik, and Timothy M. Rose

6 Sequencing of Difficult DNA Regions by SAM Sequencing 75
Keith R. Mitchelson

PART III EXPRESSION, EPIGENETIC AND COPY NUMBER ANALYSIS

7 A Global Single-Cell cDNA Amplification Method for Quantitative Microarray Analysis 91
Kazuki Kurimoto and Mitinori Saitou

8 Quantitation of MicroRNAs by Real-Time RT-qPCR 113
Caifu Chen, Ruoying Tan, Linda Wong, Richard Fekete, and Jason Halsey

9 High-Throughput SuperSAGE 135
Hideo Matsumura, Kentaro Yoshida, Shujun Luo, Detlev H. Krüger, Günter Kahl, Gary P. Schroth, and Ryohei Terauchi

10 Deep Cap Analysis of Gene Expression 147
Junpei Kurosawa, Hiromi Nishiyori, and Yoshihide Hayashizaki

11 Linking Emulsion PCR Haplotype Analysis 165
James G. Wetmur and Jia Chen

12 PAP-LMPCR: An Improved, Sequence-Selective Method for the In Vivo Analysis of Transcription Factor Occupancy and Chromatin Fine Structure 177
Richard Ingram, Arthur Riggs, and Constanze Bonifer

13 The Many Faces of MLPA 193
Thomas Ohnesorg, Erin Turbitt, and Stefan J. White

14 Assessing Gene-Specific Methylation Using HRM-Based Analysis 207
Ee Ming Wong and Alexander Dobrovic

PART IV PROBES AND ASSAYS

15 *Alu* PCR 221
Maurizio Cardelli

16 Asynchronous PCR. 231
Caifu Chen, David Ruff, and Jason Halsey

17 Novel Applications of PCR Through the Use of DNA Substrates 245
Stuart M. Wilson

18 Enhanced Solid Phase PCR for Increased Loading of Amplicon onto Solid Support 257
Daniel J. Park

19 Application of Blocking Oligonucleotides to Improve Signal-to-Noise Ratio in a PCR. 265
Hege Vestheim, Bruce E. Deagle, and Simon N. Jarman

PART V ENGINEERING AND SCREENING

20 Asymmetric Overlap Extension PCR Method for Site-Directed Mutagenesis 277
Yue-Hua Xiao and Yan Pei

21 Ribosome Display: A Technology for Selecting and Evolving Proteins from Large Libraries 283
Birgit Dreier and Andreas Plückthun

22 GLOBE: Analysis of DNA–Protein Interaction Analysis 307
Takaaki Kojima and Hideo Nakano

23 PCR DNA-Array Profiling of DNA-Binding Transcription Factor Activities in Adult Mouse Tissues 319
Yimin Sun, Jing Cheng, and Keith R. Mitchelson

24 Nucleotide Exchange and Excision Technology DNA Shuffling and Directed Evolution. 333
Janina Speck, Sabine C. Stebel, Katja M. Arndt, and Kristian M. Müller

Index *345*

Contributors

Katja M. Arndt • *Institute for Biology III, Centre for Biological Signalling Studies (bioss), Freiburg, Germany*

Constanze Bonifer • *Section of Experimental Haematology, University of Leeds, St James's Hospital, Leeds, UK*

Michael C. Borns • *Stratagene Products Division, Agilent Technologies, Inc, La Jolla, CA, USA*

Richard Boyce • *Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA*

Maurizio Cardelli • *Laboratory of Immuno-Oncology, Istituto Nazionale Riposo e Cura Anziani (I.N.R.C.A.), Ancona, Italy*

Caifu Chen • *Genomic Assays R&D, Molecular Biology Systems Division, Life Technologies, Foster City, CA, USA*

Jia Chen • *Department of Preventive Medicine, Mount Sinai School of Medicine, New York, NY, USA*

Jing Cheng • *Medical Systems Biology Research Center, Tsinghua University, Beijing, China; CapitalBio Corporation: National Engineering Research Center for Beijing Biochip Technology, Beijing, China*

Bruce E. Deagle • *Department of Biology, University of Victoria, Victoria, BC, Canada*

Birgit Dreier • *Department of Biochemistry, University of Zürich, Zürich, Switzerland*

Alexander Dobrovic • *Molecular Pathology Research and Development, Department of Pathology, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia*

Richard Fekete • *Molecular Biology Systems Division, Life Technologies, Foster City, CA, USA*

Jason Halsey • *Genomic Assays R&D, Molecular Biology Systems Division, Life Technologies, Foster City, CA, USA*

Yoshihide Hayashizaki • *RIKEN Omics Science Center, Yokohama City, Kanagawa, Japan*

Holly H. Hogrefe • *Stratagene Products Division, Agilent Technologies, Inc, La Jolla, CA, USA*

Richard Ingram • *Section of Experimental Haematology, University of Leeds, St James's Hospital, Leeds, UK*

Simon N. Jarman • *Australian Antarctic Division, Australian Marine Mammal Centre, Kingston, TAS, Australia*

Günter Kahl • *Molecular BioSciences, Biocentre, University of Frankfurt am Main, Frankfurt, Germany*

Zaheer Khan • *Genera Biosystems, Scoresby, Victoria, Australia*

Takaaki Kojima • *Laboratory of Molecular Biotechnology, Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan*
Dina Kovarik • *Northwest Association for Biomedical Research, Seattle, WA, USA*
Detlev H. Krüger • *Institute of Medical Virology, University Hospital Charité, Berlin, Germany*
Kazuki Kurimoto • *Laboratory for Mammalian Germ Cell Biology, RIKEN Center for Developmental Biology, Kobe, Japan*
Junpei Kurosawa • *RIKEN Omics Science Center, Yokohama City, Kanagawa, Japan*
Shujun Luo • *Illumina, Inc, Hayward, CA, USA*
Hideo Matsumura • *Iwate Biotechnology Research Center, Kitakami, Iwate, Japan*
Keith R. Mitchelson • *Medical Systems Biology Research Center, Tsinghua University, Beijing, China; CapitalBio Corporation: National Engineering Research Center for Beijing Biochip Technology, Beijing, China*
Kristian M. Müller • *Institute for Biology III, Centre for Biological Signalling Studies (bioss), Freiburg, Germany*
Hideo Nakano • *Laboratory of Molecular Biotechnology, Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan*
Hiromi Nishiyori • *RIKEN Omics Science Center, Yokohama City,Kanagawa, Japan*
Thomas Ohnesorg • *Molecular Development, Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, Victoria, Australia*
Daniel J. Park • *Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Parkville, Victoria, Australia*
Yan Pei • *Southwest University, Beibei, Chongqing, China*
Andreas Plückthun • *Department of Biochemistry, University of Zürich, Zürich, Switzerland*
Arthur Riggs • *Biology Division, Beckman Research Institute of City of Hope, City of Hope National Medical Center, Duarte, CA, USA*
Timothy M. Rose • *Center for Childhood Infection and Prematurity Research, Seattle Children's Research Institute, Seattle, WA, USA; Department of Pediatrics, University of Washington, Seattle, WA, USA*
David Ruff • *Genomic Assays R&D, Molecular Biology Systems Division, Life Technologies, Foster City, CA, USA*
Mitinori Saitou • *Laboratory for Mammalian Germ Cell Biology, RIKEN Center for Developmental Biology, Kobe, Japan; Department of Anatomy and Cell Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan*
Gary P. Schroth • *Illumina, Inc, Hayward, CA, USA*
Janina Speck • *Institute for Biology III, Centre for Biological Signalling Studies (bioss), Freiburg, Germany*
Jeannette P. Staheli • *Center for Childhood Infection and Prematurity Research, Seattle Children's Research Institute, Seattle, WA, USA*
Sabine C. Stebel • *Institute for Biology III, Centre for Biological Signalling Studies (bioss), Freiburg, Germany*
Yimin Sun • *CapitalBio Corporation: National Engineering Research Center for Beijing Biochip Technology, Beijing, China*
Ruoying Tan • *Molecular Biology Systems Division, Life Technologies, Foster City, CA, USA*

Ryohei Terauchi • *Iwate Biotechnology Research Center, Kitakami, Iwate, Japan*
Erin Turbitt • *Molecular Development, Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, Victoria, Australia*
Hege Vestheim • *Department of Biology, University of Bergen, Bergen, Norway*
James G. Wetmur • *Department of Microbiology, Mount Sinai School of Medicine, New York, NY, USA*
Stefan J. White • *Molecular Development, Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, Victoria, Australia*
Stuart M. Wilson • *Microsens Medtech Ltd, London, UK*
Ee Ming Wong • *Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia; Molecular Pathology Research and Development, Department of Pathology, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia*
Linda Wong • *Molecular Biology Systems Division, Life Technologies, Foster City, CA, USA*
Yue-Hua Xiao • *Southwest University, Beibei, Chongqing, China*
Bin Yin • *Cyrus Tang Hematology Center, Jiangsu Institute of Hematology, First Affiliated Hospital, Soochow University, Suzhou, China*
Kentaro Yoshida • *Iwate Biotechnology Research Center, Kitakami, Iwate, Japan*

Chapter 1

Setup of a PCR Laboratory

Zaheer Khan

Abstract

PCR represents an extremely powerful and central molecular biology method. At the heart of its power is the exquisite sensitivity offered: single molecule detection in certain contexts. However, with great power comes great responsibility. Contamination of reagents or test samples with amplifiable material, such as previous reaction products, can be crippling to scientists applying PCR protocols. Prevention of PCR contamination is far and away preferred over eradication. This chapter sets out to offer guidance as to how to use PCR while minimising contamination problems.

Key words: PCR laboratory setup, PCR contamination, PCR reagents, PCR equipment

1. Introduction

From its beginnings in 1985, PCR technology has become a cornerstone of molecular biology. As PCR became more widely used, scientists discovered its strengths and weaknesses. Amplification to detectable levels can be achieved from even single copy of template nucleic acids. PCR can be used to amplify from not only DNA but also RNA and the products can now be detected in real time. However, the attribute that makes it most appealing is the same attribute that is the source of much frustration. PCR contamination, once established, can be very difficult to rid oneself of. The best approach is to prevent PCR contamination from happening in the first instance.

With a good laboratory setup and good technique, many of the problems associated with PCR can be avoided. However, one must always remember that all we can do is to reduce the risk to acceptable levels. The more procedures we implement, the further reduced are the risks of PCR contamination. This must be

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_1, © Springer Science+Business Media, LLC 2011

balanced with creating a work environment that is comfortable to work in. Implementing too many additional procedures may result in a tendency to not follow what are perceived as unnecessary procedures.

The appropriate balance between ease of work and preventing PCR contamination will be different for different organisations. A laboratory manufacturing PCR diagnostics kits to regulatory standards would need to take the most stringent measures to prevent PCR contamination, whereas a laboratory which performs PCR infrequently for research purposes may not need such stringent controls, and the balance would be tipped more toward ease of work. These are decisions which, ultimately, each laboratory must make based on their requirements.

This chapter outlines some general considerations for setting up a PCR laboratory with a view to preventing contamination.

2. Materials

Materials should be stored as specified by the manufacturer. Where a material is made within the laboratory, all bottles should be clearly and uniquely labelled with at least the manufacturer's initials or name, name of the solution or the material, and the expiry date. Materials should be placed in the rooms they are to be used in, and where one material is required in multiple rooms, each room should have a dedicated stock. Materials should not be moved from higher copy to lower copy rooms as this can increase the risk of transferring contamination between rooms. If necessary, but as a last resort, procedures should be implemented to decontaminate materials (see Note 1). In case of laboratory contamination, some guidelines are provided in Subheading 3.

There are many different pieces of equipment which are important for a PCR laboratory. Later, the placement of critical equipment is specified, but here some of the more important pieces of equipment are introduced.

1. Fridges and freezers are used for storing reagents and samples. Generally, a fridge or freezer should display the temperature or contain probes for monitoring temperature. Temperature should be monitored regularly such that any variations of temperature outside specification can be recorded. Preferably an alarm should be connected. It is also important to know for how long a fridge or freezer has been out of specification. Freezers are typically available as either −20 or −80°C models. A −80°C freezer is recommended for long-term storage of nucleic acids and a −20°C freezer is adequate for shorter periods of storage and is generally used to store PCR reagents in common use (1).

Ideally, calibration should be performed annually. Frost-free freezers should be avoided for the storage of molecular biology reagents (see Note 2).

2. Pipettes are fundamental tools within the laboratory. They come as single channel, multichannel, and/or stepper/repeater pipettes. They are available as positive-displacement or air-displacement models. Positive-displacement pipettes are more accurate for pipetting viscous liquids (2) and when used appropriately can help to prevent contamination as the entire tip and mechanism is ejected after use. Use of barrier tips with air-displacement pipettes is strongly recommended to reduce contamination derived from PCR products or other sources via aerosols (3). Calibration of pipettes is critical to ensure accurate and reproducible results. Currently, there are no definitive regulations on how often calibration should occur. However, the *Center for Drug Evaluation and Research, FDA, Guidance for Industry, Subpart D, Equipment states* "*the key point is that the calibration schedule should be frequent enough to assure data validity...*". Factors which affect how often calibration should occur include:

   - The liquid that is dispensed. Corrosive liquids can form vapours which can damage the internal mechanisms of the pipette.
   - Handling of pipettes. Pipettes that are often dropped are more subject to wear and tear, and will require more frequent calibrations.
   - The accuracy required. Facilities that require high accuracy in their procedures need to regularly monitor the adequacy of their pipettes.

   In general terms, the author recommends that most research laboratories should calibrate pipettes every 6 months. Facilities which operate under GMP regulations should consider calibration every 3 months.

3. Thermocyclers are required to perform PCR. There are many manufacturers making a variety of models. Thermocyclers include machines with 96- or 384-well format blocks. Some have interchangeable blocks. Reactions can occur in single tube, strips, 96- or 384-well plates, or capillaries. Reaction vessel volumes for use in conventional PCR are typically 0.2 or 0.5 mL. Ramp rates and uniform heat distribution are important factors influencing the consistency of operation. It is recommended that these be calibrated on a regular basis. When utilising several different brands of PCR machines, one should not assume that they will perform identically. Particular attention should be paid to the ramp rates, tube size, and block

architecture. Real-time PCR technology allows amplification and analysis to take place simultaneously without the need for further handling. This reduces PCR contamination risk as the completed reaction can be discarded immediately after processing without exposure to the laboratory environment.

4. Cleaning with 2–6% hypochlorite (w/v) has been shown to degrade DNA (4) (see Note 3). At this concentration, the levels of hypochlorite are similar to household bleach; always refer to the manufacturer's MSDS for safety and disposal instructions. Regular cleaning of surface area before and after work can reduce the spread of DNA and prevent contamination. A value of 70% (v/v) ethanol is often used to wipe benches after hypochlorite treatment to prevent it from interfering with PCRs. Regularly changed bench liners and a regular laundry schedule are important elements.
5. Planning a PCR laboratory: the most common setup for a PCR laboratory involves three separate areas. It is highly recommended that at least three separate rooms should be used for the three major steps of PCR: the Master Mix preparation and pre-PCR reagent handling and addition, the sample addition, and the PCR and analysis. Implementing such physical separation between the three steps reduces the risk of products contaminating reagents or samples to be included in future PCRs.

These areas are generally termed as a Pre-PCR Room, a Low Copy Room, and a High Copy Room. Work flow should occur in a unidirectional manner (Fig. 1), and at no point should samples

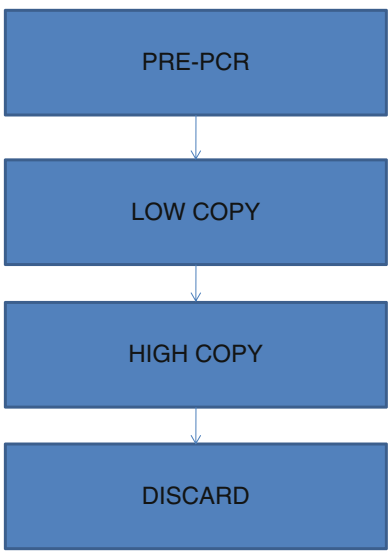


Fig. 1. Recommended workflow through a PCR laboratory. Workflow should always be initiated in the pre-PCR area and flow progressively through the other work areas. The flow is unidirectional and materials should never flow backwards.

or equipment from downstream, including laboratory coats and gloves, be transferred to upstream areas. Each room should have designated equipment and instruments which should not be moved between rooms. Ideally, the Pre-PCR Room and Low Copy Room should be positively pressurised relative to the corridor. The High Copy Room should be negatively pressurised relative to the corridor. In this manner, PCR aerosols or contaminated dust are discouraged from entering the Low Copy and Pre-PCR areas (see Note 4). Where such physical separation is not possible, three separate areas should be established. If only two rooms are available, the low copy area and the Pre-PCR area can be shared. Separate equipment and instruments should be arranged for each area and transfer of equipment and instruments between areas should be avoided. For the pre-PCR setup, a laminar flow hood should be used.

3. Methods

1. Routinely one must gown up and glove up upon entering the room and degown and remove gloves just before leaving the room (see Note 5). This is good practise for all laboratory areas and ensures that whatever work has been done in that area stays in that area and does not travel to other areas.
2. Use of bench liners will contain any spills that may interfere with the PCR; these should be regularly changed and monitored. Any spills should generally be wiped using absorbent disposable wipes. When wiping spills, do not smear the surface, but absorb the spill and dispose of the wipe.
3. Regular cleaning of surfaces with 2–6% hypochlorite (w/v), followed by 70% (v/v) ethanol ensures a good clean environment to work in (5). Along with maintaining a clean environment, one must always be conscious of spreading contamination. Contaminated bench tops and door handles are major risks (see Note 6). Regular changing of gloves can prevent such spread. It is further recommended that gloves be removed immediately prior to exiting a room (see Subheading 3, step 1); gloved hands should not come into contact with door handles as this can easily spread contamination between laboratory areas.
4. Always centrifuge samples before opening as these can be a source of aerosols which can contaminate laboratory areas.
5. Gel areas are a potent source of contamination. Gel areas must always be located in the High Copy area and where possible should be segregated from the rest of the laboratory.

Particular care must be taken with buffers such as those used during electrophoresis, as PCR products have been in direct contact with the solutions. Always change gloves before leaving the gel area.

6. All equipment should be dedicated for each area and should not be moved between areas. Equipment should be regularly monitored for performance and calibrated as required.
7. The Pre-PCR Room is routinely used for preparation of Master Mixes and storage of pre-PCR reagents. It is critical that this area be kept free of any DNA or RNA, especially those related to work routinely performed in the laboratory. Any contamination here can be hard to identify, time-consuming to investigate, and expensive to remedy. General guidelines on how to deal with established PCR contamination are included later in this chapter. One must always be very cautious when bringing material into the room. The Pre-PCR Room should contain everything that is required within the room such that there is no need for laboratory personnel to exit and re-enter the room during pre-PCR processing. The Pre-PCR Room should contain a –20 or –80°C freezer for storage of PCR reagents and a vortex for mixing reagents; however, enzymes or enzyme mixes should not be vortexed. The Pre-PCR Room should also contain a microfuge to "pulse down" reagents following mixing (see Note 7). Always check to ensure no liquid is present in the lid as this can drop out and contaminate bench surfaces and gloves. Bench surfaces should be regularly cleaned before and after setup using 2–6% hypochlorite (w/v), followed by 70% (v/v) ethanol. Dedicated pipettes and tips for liquid handling should be used. A laminar flow cabinet is highly recommended for preparing Master Mixes. If laminar flow cabinets cannot be utilised, special PCR cabinets can be used. Many cabinets come equipped with UV lights (see Note 8). These should be turned on before and after PCR setup for at least 30 min. UV radiation can also be used to remove contaminations from PCR Master Mixes for use in the PCR. Studies have shown that exposure to UV light for a period of 30 min can inactivate many DNA species in solution by approximately five orders of magnitude (6) (see Note 9). Other equipment used in preparation can also be routinely decontaminated using UV light. It is suggested that this be done in conjunction with bleach and ethanol cleaning on a regular basis as part of the laboratory cleaning schedule.
8. The Low Copy Room is often used for the preparation, storage, and addition to reaction mixtures of DNA or RNA samples used in PCR. No PCR product or DNA clones containing target sequences should be handled in this area. This area

should contain a –20 or –80°C freezer for storage of DNA, vortex mixers, and centrifuges. Consideration should be given to how much freezer space is required; in some cases, the laboratory may desire to store samples for a number of years, in which case provisions should be made to ensure there is adequate space for storage (see Note 10). Samples such as blood, serum, or cultured cells can be brought into this room for the purpose of extracting DNA or RNA, assuming appropriate safety precautions are in place. If working with infectious materials, ensure that appropriate systems and equipment are in place for the protection of both laboratory staff and samples (see Note 11). It is ultimately the responsibility of the laboratory head and senior staff to comply with any biosafety regulations applicable to their laboratory. Dedicated tools and equipment should be utilised in this room. Use of barrier tips or positive-displacement pipettes discourages the spread of aerosols and should be used throughout the PCR procedure. To further reduce the spread of aerosols and liquid collected around vessel lids, one should always centrifuge samples briefly before opening. Use of screw cap tubes is preferred over "popping" lids as the latter can increase the probability of aerosol formation. When extracting DNA or RNA, a method that requires the least amount of handling is recommended. This not only decreases the risk of spreading genetic material, but also reduces the risk of contamination between samples. For large numbers of extractions, automated systems provide a secure, contained environment where DNA or RNA extractions can occur with a reduced risk of spreading material throughout the laboratory. The closed space allows for quick and easy decontamination between runs. Automated DNA extraction systems provide reproducible and reliable quantities and qualities of DNA which are not subject to the variation which can occur when human operators are involved. They reduce the possibility of sample mix-up which is a very important consideration for diagnostic laboratories. A disadvantage of these systems can be their reduced sensitivity with some samples (7).

9. The High Copy Room is so termed because of the presence of high copies of DNA and PCR products. This area is used for the amplification of low copies of DNA to yield high copies of DNA. A PCR mix (post-template addition) is brought into this area and amplified using a thermocycler. The products generated by PCR are a major source of contamination within the laboratory. This contamination principally occurs from aerosols which arise from the process of pipetting and manipulating the DNA, contaminated liquids being splashed, and surface-to-surface contact. Several strategies have been

introduced to minimise this contamination; the same methods can be utilised in this area. These contaminants cause few problems if confined to the High Copy Room but can cause havoc if allowed to spread to the other areas. It is therefore critical that no mixing of reagents and equipment be allowed between this area and the other areas of the laboratory. This area contains much of the equipment, such as thermocyclers, centrifuges, qPCR facilities, sequencing facilities, electrophoresis equipment, and imaging equipment including transilluminators and cameras for photographing gels. It also often serves as the storage site for PCR products. Although PCR products can be stored in this area they must not be taken out of this area unless being disposed of in a suitably contained manner.

10. Discard: all items for disposal in the High Copy Room are transferred only when appropriately contained to prevent exposure of other areas of the laboratory to high copy material. Waste contract services are recommended as they will supply and control the management of waste generated by the laboratory. There are three major classifications of hazardous waste (8) routinely generated during normal PCR laboratory functioning. The first is infectious waste/highly infectious waste which is routinely used for the disposal of PCR products, contaminated agar plates, live cultures, human cells and blood, and disposables that have been in contact with the above. The second is sharps which includes objects or devices that have acute, rigid corners, edges, points or protuberances capable of cutting or penetrating the skin, e.g. hypodermic needles, glass, scalpel blades, and lancets. All sharps are hazardous because of the potential to cause cuts and punctures. Sharps may also be contaminated with toxic, infectious, or radioactive materials, which would increase substantially their potential to cause harm. Thirdly, there is hazardous chemical waste. Chemical waste consists of discarded chemicals (solid, liquid, or gaseous) that are generated during disinfecting procedures or cleaning processes, for example. They may be hazardous (toxic, corrosive, and flammable) and must be used and disposed of according to the specification formulated on each container. Nevertheless non-explosive residues or small quantities of outdated products may be treated together with infectious waste.
11. In cases where the above measures relating to physical separation are impossible, or as an additional measure, "temporal" separation of areas can be employed. This separates work tasks in time such that low copy and high copy materials are never handled at the same time. Combined with appropriate cleaning schedules and the use of separate equipment, consumables,

and areas, this approach can help to reduce the risk of introducing PCR contamination into reagent stocks or template material (see Note 12).

12. Enzymatic treatment of PCR Master Mix is an elegant, optional measure to help to limit "PCR carryover" (9). However, this should not be viewed as a replacement for any of the other measures discussed in this chapter (see Note 13). The basic principle involves the substitution of dTTP with dUTP, where dUTP is used as substrate during polymerisation catalysed by compatible polymerases. The PCR product formed after this substitution is susceptible to enzymatic cleavage. Incubation of PCR Mix with a quantity of uracil-DNA *N*-glycosylase before addition of any template removes PCR product "carry over" that may have found its way into the Master Mix. Subsequent heating at 90°C destroys the enzyme. Many kits are available to streamline this process and in order for this method to be effective it must be incorporated as early as possible. UV irradiation of the PCR Master Mix is another option which can be used to limit PCR carryover (10) (see Note 9).
13. Negative controls must always be employed so that any contamination can be detected as early as possible. When contamination is found, a pre-planned strategy must be put into action as quickly as possible.
14. If PCR contamination is detected in the laboratory, one of two approaches can be used. The first approach centres on the quick eradication of PCR contamination. This approach will quickly remove the source of the contamination but can be very expensive. It will not identify the source of the contamination. In this approach, the laboratory must discard all reagents used in PCR preparation, including primers, buffers, dNTPs, water, and enzymes. Also included can be the discarding of equipment and consumables such as pipettes and PCR plates or tubes. Thorough decontamination of workspaces and any equipment that cannot be discarded should be performed. The second approach involves the testing of all reagents and equipment. This can be a very time-consuming exercise, but the source of the contamination can be isolated. It can be less expensive (depending on the resources consumed in terms of personnel-time, reagents consumed, and ultimate success of the process) as only the source of the contamination need be discarded and measures can be put in place to prevent future contaminations of the identified component. In this approach, fresh reagents are prepared from known PCR-clean sources. Each suspected reagent is tested in isolation to identify which component is the source of the contamination.

15. Although the utilisation of the aforementioned techniques will never ensure the prevention of PCR contamination, they will reduce the risk of PCR contamination. By thinking of all possibilities and planning the work environment, we can substantially reduce the risk of contamination occurring.
16. Good training of laboratory staff is essential regardless of how well planned a laboratory is. Staff should understand the reasons for the systems that are in place and the cost of having PCR contamination within the laboratory (see Note 14). Formal laboratory training of all staffs in laboratory protocols and procedures should be accompanied by signed training forms and regular reviews to ensure compliance. The laboratory staff will ultimately determine how the laboratory functions, and as such, the resources required for them to be able to perform their duties efficiently and effectively should be made available to them.

4. Notes

1. Decontamination should be viewed as a last resort. Prevention of PCR contamination is by far the preferred alternative.
2. Frost-free freezers cycle through rounds of relatively warmer and cooler temperature cycles. This results in multiple freeze–thaw events which can be damaging to the stored PCR reagent or template nucleic acid.
3. Hypochlorite bleach must be prepared relatively fresh and be within specification. Active species degrade over time to result in reduced bleaching potency.
4. Installation of appropriately pressurised areas is easier during the "fit-out" stage of the setting up of a laboratory suite. Retrofitting the relevant rooms is possible but will likely prove much more problematic from an engineering and installation perspective, and as such, will likely prove much more expensive.
5. Dedicated separate gowns are used in separate areas. Gowning/degowning and gloving up and removing gloves must be seen as only the beginning in terms of how these items of protective-wear can help to minimise the risk of PCR contamination. How these are used, what they touch, and when they are replaced all determine how effective their use is. In particular, gloves should be changed frequently, especially when a high risk or suspected surface has been contacted.
6. Surface-to-surface contact represents one of the highest risks with respect to transmission of PCR contamination. The same principles, which apply in aseptic technique used in cell

culture or microbiology in terms of avoiding "potentially contaminated" surface contacts, apply in the handling of materials in the PCR laboratory. Any surface-to-surface contact (including gloves – see Note 5) should be viewed as a potential contamination transfer event.

7. In particular, aerosols and wet surfaces facilitate undesirable transfer of PCR contaminants. Centrifugation of reagent or specimen vessels to remove such agents from around the lids of vessels can help to prevent PCR contamination.
8. UV light is very damaging to the eyes and skin. Operators should ensure that they always wear appropriate protective-wear to protect themselves while working with UV light. Operators should also ensure that other users of the workspace are appropriately protected from any potential exposure.
9. If UV light is used to treat Master Mixes for the purpose of decontamination, it should be used in the absence of dNTPs or enzymes or exposure should be limited in their presence as these are also damaged by UV light exposure. Decontamination in this manner should be seen as a last resort since prevention of PCR contamination is infinitely preferable to its attempted removal (see Note 1).
10. Freezer storage redundancy should be designed into the operation of the laboratory in each area. This will allow provision of "spill over room" (free space available for the transfer of reagents or samples from one freezer to another) for emergencies and for routine freezer defrost and cleaning activities.
11. This chapter focuses on handling with particular regard to preventing PCR contamination. Hazards such as infectious agents handling must also be considered when planning the laboratory suite and training personnel, such that containment facilities, procedures, and certification are in place to handle specimens with the appropriate levels of control, containment, and operator safety.
12. Temporal separation should be viewed as a last resort replacement for physical separation of PCR areas. Wherever possible, separate equipment and consumables, and rigorous cleaning schedules should be employed.
13. Enzymatic control of "carry over" contamination should not be viewed as a replacement for the other measures for PCR contamination control set out in this chapter. This should be considered as an additional measure. It should be noted that certain polymerases in common use are not compatible with this approach due to the requirement for dUTP to be incorporated efficiently during polymerisation.
14. Staff or student laboratory orientation and training are so often treated without the deserved level of importance. In the context

of staffing a PCR facility, task-appropriate staff training is imperative. One inexperienced and/or inappropriately trained operator can inadvertently cause enormous damage in terms of down-time, investigational, corrective and preventative effort, and expense. This is of the utmost importance in cases where dangerous infectious agents or hazardous chemicals are handled during laboratory procedures (see Note 11).

References

1. Halfon, P., Khiri, H., Gerolami, V., Bourliere, M., Feryn, J.M., Reynier, P., Gauthier, A., and Cartouzou, G. (1996) Impact of various handling and storage conditions on quantitative detection of hepatitis C virus RNA. *J. Hepatol.* **25**, 307–11.
2. Anderson, S.C. and Cockayne, S. (2003) Clinical Chemistry: Concepts and Applications. McGraw-Hill, New York.
3. Wilson, C.L. (ed.) (2007) Microbial Food Contamination, Second Edition. CRC, Boca Raton.
4. Kemp, B.M. and Smith, D.G. (2005) Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic Sci. Int.* **154**, 53–61.
5. Prince, A.M. and Andrus, L. (1992) PCR: how to kill unwanted DNA. *Biotechniques* **12**, 358–60.
6. Sarkar, G. and Sommer, S.S. (1990) Shedding light on PCR contamination. *Nature* **343** (6253), 27.
7. Schuurman, T., van Breda, A., de Boer, R., Kooistra-Smid, M., Beld, M., Savelkoul, P., and Boom, R. (2005) Reduced PCR sensitivity due to impaired DNA recovery with the MagNA Pure LC total nucleic acid isolation kit. *J. Clin. Microbiol.* **43**, 4616–22.
8. The 10 categories of HCRW WHO Healthcare waste management (HCWM). (Accessed January 25, 2010, at http://www.healthcarewaste.org/en/128_hcw_categ.html).
9. Longo, M.C., Berninger, M.S., and Hartley, J.L. (1990) Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene* **93**, 125–8.
10. Niederhauser, C., Höfelein, C., Wegmüller, B., Lüthy, J., and Candrian, U. (1994) Reliability of PCR decontamination systems. *PCR Methods Appl.* **4**, 117–23.

Chapter 2

Long-Range PCR with a DNA Polymerase Fusion

Holly H. Hogrefe and Michael C. Borns

Abstract

Proofreading DNA polymerase fusions offer several advantages for long-range PCR, including faster run times and higher fidelity compared with *Taq*-based enzymes. However, their use so far has been limited to amplification of small to mid-range targets. In this article, we present a modified protocol for using a DNA polymerase fusion to amplify genomic targets exceeding 20 kb in length. This procedure overcomes several limitations of *Taq* blends, which up until recently, were the only option for long-range PCR. With a proofreading DNA polymerase fusion, high-molecular-weight amplicon can be generated and analyzed in a single day, and a significant proportion is expected to be error-free.

Key words: Long PCR, Long-range PCR, Fusion DNA polymerase, Uracil

1. Introduction

Long-range PCR is used to prepare specific high-molecular-weight DNA fragments for a variety of applications, including cloning, genome mapping and sequencing, and contig construction. Routine amplification of large genomic targets remains challenging, despite the availability of specialized long-range PCR enzymes. The majority of researchers use *Taq* blended with a small amount of proofreading enzyme (*Taq* blend) to amplify targets greater than 10 kb in length (1). Such blends, however, exhibit several disadvantages, including: (1) low fidelity which leads to a high percentage of error-containing clones (e.g., 100% of 10 kb fragments are expected to contain at least one error; Table 1); and (2) long PCR extension times that can damage DNA and lead to run times exceeding 12 h.

In this chapter, we describe an improved method for long-range PCR that provides faster, more accurate amplification of large genomic targets. The protocol features a proofreading DNA

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_2, © Springer Science+Business Media, LLC 2011

Table 1
Comparison of long-range PCR enzymes

Long PCR enzyme	Error rate	% Clones with errors[a]			Extension time	Overall run time (h)[b]		
		1 kb	10 kb	20 kb		1 kb	10 kb	20 kb
Taq blend	$8–14 \times 10^{-6}$ (6)	16–28	100	100	60 s/kb	1.3–1.6	6.5–6.8	11.5–12.5
Pfu fusion	1.3×10^{-6} (5)	3	26	52	30 s/kb	1	3.2	6.5

[a]Calculated as described (11) assuming a duplication of 20 (10^6-fold amplification)
[b]Thermocycle programmed according to parameters recommended here (*Pfu* fusion) or in manuals of three different long PCR *Taq* blends

polymerase fusion (2) rather than a *Taq* blend, a modified cycling protocol designed to minimize PCR run time, and a single PCR additive (DMSO) to facilitate replication of difficult GC-rich sequences. In the example provided, we use Herculase II fusion DNA polymerase (Herculase II), which consists of *Pfu* DNA polymerase fused to a double-stranded DNA binding protein (*Pfu* fusion) and a thermostable dUTPase. Compared with wild-type enzymes, *Pfu* fusion exhibits higher processivity (185 bases versus 10–15 bases for *Pfu* or 10–42 bases for *Taq* (3, 4)), which allows researchers to use shorter extension times and perform long-range PCR in half the time required for *Taq* blends (Table 1). In addition to speed, the *Pfu*-only Herculase II enzyme provides superior fidelity compared with *Taq* blends (4–6), ensuring that a significant proportion of amplified DNA is error-free (Table 1).

Herculase II is recommended over other DNA polymerase fusions because the formulation includes a thermostable dUTPase that enhances length capability. Decxyuracil is a potent inhibitor of archaeal DNA polymerases (e.g., *Pfu*, Vent, and DeepVent) that arises during PCR by deamination of dCTP to dUTP, and subsequent incorporation of dUTP into amplicon (7). Archaeal DNA polymerases possess a unique "read-ahead" function, and stall when decxyuracil is encountered in the template strand (8). Uracil poisoning of archaeal DNA polymerases (alone or in *Taq* blends) has significant implications for long-range PCR due to the use of long extension times. Prolonged exposure of nucleotides to heat leads to increased formation of dUTP, incorporation of uracil in early PCR cycles, and reduced amplification efficiency (9). In the Herculase II formulation, dUTPase is added to eliminate dUTP by conversion to dUMP and PP_i. In the absence of dUTPase, all long-range PCR enzymes suffer from uracil poisoning, which can be largely overcome in *Taq* blends by reducing the proportion of proofreading enzyme, at a cost to fidelity (error rate comparable to *Taq* alone (5)).

The protocol provided below has been used to amplify genomic targets up to 23 kb in length. We have included instructions

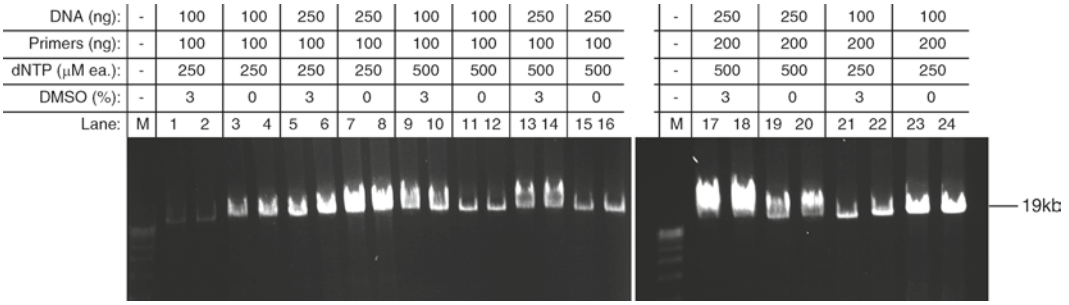


Fig. 1. Optimization parameters for long PCR with a DNA polymerase fusion. An 18.4 kb β-globin fragment was amplified in duplicate reactions from human genomic DNA (Promega G3041) using recommended reaction conditions and cycling parameters (*see* Subheading 3). Adjustments in DNA template, primer, nucleotide, and DMSO concentrations are indicated. Primer sequences are: (Forward) 5′ CACAAGGGCTACTGGTTGCCGATT and (Reverse) 5′ CCTGCATTTGTGGGGTGAATTCCTTGCC. In this example, highest product yields were achieved under two different sets of conditions: (lanes 7–8) 250 ng DNA, 100 ng each primer, 250 μM each dNTP, and no DMSO; and (lanes 17–18) 250 ng DNA, 200 ng each primer, 500 μM each dNTP, and 3% DMSO.

for amplification of a broad range of target sizes, including both mid-range (1–10 kb) and long (>10 kb) DNA fragments. We provide tips for isolating suitable quality genomic DNA, as well as PCR primer sequences for an 18.4 kb β-globin fragment that can be used to optimize long-range PCR procedures in individual laboratories (Fig. 1).

2. Materials

1. *PCR components.* Commercial long-range PCR enzymes typically come with reaction buffer and one or more enhancers or additives to facilitate amplification of GC-rich target sequences. Herculase II fusion DNA polymerase (Agilent Technologies – Stratagene Products) includes 5× reaction buffer (provides a final 1× Mg^{2+} concentration of 2 mM) and dimethylsulfoxide (DMSO). Researchers need only provide deoxynucleotides, primers, and DNA template of sufficient quality for long-range PCR.
2. *DNA template.* The most critical component of long-range PCR is the DNA template. Great care should be exercised to isolate intact, high-molecular-weight (>50 kb) genomic DNA. Suitable quality templates are routinely isolated using the DNA Extraction Kit or RecoverEase DNA Isolation Kit (Agilent Technologies – Stratagene Products). Potential shearing of genomic DNA template is minimized by using wide-bore tips for pipetting or mixing of the template. High-molecular-weight templates should be stored at 4°C, and not frozen.

3. *PCR primers.* Primers for long-range PCR should be at least 23 bp in length, and ideally gel- or HPLC-purified. Primer pairs should be designed with balanced melting temperatures (T_m) of at least 60°C. The resulting high-annealing temperature promotes specificity and discourages secondary structure formation. Standard primer design rules should be applied, including analyzing primer sequences for potential duplex and hairpin formation as well as false priming sites, to ensure the highest yield of specific PCR products.
4. *Thermocycling and analysis of PCR products.* Cycling conditions for long- range PCR (Subheading 3.2) were developed for single-block temperature cyclers, including the DNA Engine PTC-200 (BioRad), and GeneAmp PCR System 9700 (ABI). Cycling parameters are not necessarily transferable between thermal cyclers designed by different manufacturers, and although any thermocycler is suitable for long-range PCR, further optimization may be required for other instruments. After cycling, PCR products up to 20 kb in length can be analyzed by gel electrophoresis using 0.6–0.7% agarose gels, stained with ethidium bromide and visualized on a suitable UV imaging system.

3. Methods

The following protocol is recommended for amplifying targets greater than 10 kb in length. See Note 1 to modify protocol for amplification of targets <10 kb.

3.1. PCR Set-Up

Prepare 50 µl PCRs in sterile thin-walled PCR tubes. Add reagents in the order listed (see Note 2). Vortex gently.

Distilled water	Xµl to a final volume of 50 µl
5× Herculase II reaction buffer	10 µl
dNTP mix (25 mM each dNTP)	0.5 µl
DNA template	250 ng genomic DNA or 15 ng lambda/plasmid DNA
Forward primer	200 ng
Reverse primer	200 ng
Herculase II enzyme	1 µl
DMSO	0–4 µl (titrate over a range of 0–8%)

3.2. PCR Cycling

If the temperature cycler lacks a heated cover, or if extension times are >15 min, overlay each reaction mixture with 50 µl mineral oil. Perform PCR using the cycling conditions provided

below, which were developed for single-block temperature cyclers (see Note 3 and Subheading 2, item 4).

Initial denaturation	92°C, 2 min
10 cycles	92°C, 20 s Primer T_m – 5°C, 20 s 68°C, 30 s per kilobase of PCR target
20 cycles	92°C, 20 s Primer T_m – 5°C, 20 s 68°C, 30 s per kilobase of PCR target plus 20 s per cycle
1 cycle	68°C, 8 min

3.3. Analysis of PCR Products

PCR products are analyzed by gel electrophoresis using an appropriate percentage agarose gel. According to Maniatis, the effective range of separation of linear DNA molecules is 0.8–10 kb for a 0.7% agarose gel, and 1–20 kb for a 0.6% agarose gel (10). For maximum separation and resolution, we recommend pulse field gel electrophoresis with a 1.0% agarose gel.

4. Notes

1. *Modified protocol for shorter targets.* Herculase II can also be used to amplify targets less than 10 kb in length. Reactions should be set up according to the following protocol:

Distilled water	Xµl to a final volume of 50 µl
5× Herculase II reaction buffer	10 µl
dNTP mix (25 mM each dNTP)	0.5 µl
DNA template	100 ng genomic DNA or 1 ng lambda/plasmid DNA
Forward primer	100 ng
Reverse primer	100 ng
Herculase II enzyme	0.5 µl (<1 kb targets) or 1 µl (>1 kb targets)

Cycle shorter targets as follows:

Initial denaturation	95°C, 2 min
30 cycles	95°C, 20 s Primer T_m – 5°C, 20 s 72°C, 30 s (<1 kb) or 30 s per kilobase (1–10 kb) of PCR target
1 cycle	72°C, 5 min

2. *Optimization of reaction components for long-range PCR.* All PCR amplification reactions require optimization to achieve

the highest product yield and specificity. Critical optimization parameters for successful amplification of long targets are outlined below, in order of priority. As shown in Fig. 1, optimization of DNA template, primer, and DMSO concentrations can have a significant impact on PCR product yield.

(a) *DNA template*. Successful amplification of long targets is dependent on the purity, integrity, molecular weight, and concentration of the DNA template. As discussed in Subheading 2, item 2 above, great care should be taken to isolate intact, high-molecular-weight (>50 kb) genomic DNA. Yields are generally improved by increasing the amount of genomic DNA, although excess DNA template can be inhibitory. We recommend titrating genomic DNA over the range of 150–400 ng for targets greater than 10 kb in length. When amplifying low-complexity targets (e.g., lambda or plasmid DNA), input DNA should be titrated from 15 to 60 ng in a 50 µl reaction volume.

(b) *Primers*. PCR primers should be purchased from a reputable vendor, and for best results, gel- or HPLC-purified. Product yield is generally improved by adjusting the ratio of primer versus template. As a starting point, we recommend using 0.5 µM each primer for targets >10 kb, which is equivalent to approximately 200 ng of a 25 base oligonucleotide primer in a 50 µl reaction volume.

(c) *DMSO*. DMSO facilitates amplification of extra-long or GC-rich targets by destabilizing secondary structures in the DNA template that impede polymerization or primer annealing. DMSO may increase DNA polymerase error rate slightly (<50% increase with 3% DMSO), so its use should be avoided in cases where there is no benefit to yield or specificity. The DMSO concentration should be titrated for each primer-template system, as the degree of improvement will vary according to target length, complexity, and GC content. As a guideline, we recommend titrating DMSO in 1% increments between 0 and 3% for targets up to 20 kb, and between 3 and 6% for targets >20 kb. DMSO concentrations up to 8% may enhance amplification of targets containing GC-rich sequences. DMSO reduces primer T_m, and reoptimization of PCR annealing temperature may be required for maximum yield.

(d) *Nucleotides and magnesium*. Nucleotide concentrations of 250 µM each are sufficient for amplifying targets up to 20 kb in length, and further optimization is not usually required. The Herculase II reaction buffer provides the magnesium ion concentration that is optimal for the enzyme (final 1× Mg^{2+} concentration of 2 mM). Increasing

the magnesium concentration may lower fidelity (11) or specificity, and should be avoided.

3. *Optimization of cycling conditions for long-range PCR*. Key modifications to standard protocols include lowering the denaturation (from 95 to 92°C) and extension (from 72 to 68°C) temperatures to minimize thermal damage to the DNA template. With DNA polymerase fusions, short extension times of 30 s per kb are maintained for the first ten cycles, after which 20 s are added to the total extension time at each of cycles 11–30 (e.g., cycle 11, 30 s/kb + 20 s; cycle 12, 30 s/kb + 40 s, and so on). Modifying extension parameters in this fashion provides significant improvements in the yield of long targets, while overall run times are increased by not more than 25%.

Cycling parameters provided in Subheading 3.2 may require further optimization depending on the primer-template set and thermal cycler used. For example, raising or lowering the annealing temperature may provide further improvements in specificity or yield, respectively. GC-rich targets, which may be difficult to melt, typically require the use of more stringent denaturation conditions (e.g., 98°C for 40 s) and higher extension temperatures (72°C). Finally, the duration of the denaturation, annealing, and extension steps may require further adjustment depending on the ramp rate of the thermal cycler used.

References

1. Barnes, W.M. (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proc. Natl. Acad. Sci. USA* **91**, 2216–20.
2. Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M., and Vander Horn, P.B. (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Res.* **32**, 1197–207.
3. Hogrefe, H.H., Cline, J., Lovejoy, A., and Nielson, K.B. (2001) DNA Polymerases from Hyperthermophiles. *Methods Enzymol.* **343**, 91–116.
4. (2006) Herculase II fusion DNA polymerase product literature. *Strateg. Newsl.* **19**, 4–5, 34–35.
5. Hogrefe, H.H. and Borns, M.C. (2003) High Fidelity PCR Enzymes, *PCR Primer*, Dieffenbach, C.W., and Dveksler, G.S., eds., Cold Spring Harbor Laboratory Press, Plainview, NY.
6. Borns, M. and Hogrefe, H. (2000) Unique DNA polymerase formulation excels in a broad range of PCR applications. *Strateg. Newsl.* **13**, 1–3.
7. Hogrefe, H.H., Hansen, C.J., Scott, B.R., and Nielson, K.B. (2002) Archaeal dUTPase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation. *Proc. Natl. Acad. Sci. USA* **99**, 596–601.
8. Greagg, M.A., Fogg, M.J., Panayotou, G., Evans, S.J., Connolly, B.A., and Pearl, L.H. (1999) A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. *Proc. Natl. Acad. Sci. USA* **96**, 9045–50.
9. Arezi, B., Xing, W., Sorge, J.A., and Hogrefe, H.H. (2003) Amplification efficiency of thermostable DNA polymerases. *Anal. Biochem.* **321**, 226–35.
10. Maniatis, T., Fritsch, E.F., and Sambrook, J. (eds.) (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, NY.
11. Cline, J., Braman, J.C., and Hogrefe, H.H. (1996) PCR fidelity of *Pfu* DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.* **24**, 3546–51.

Chapter 3

Isolation of Genomic Insertion Sites of Proviruses Using Splinkerette-PCR-Based Procedures

Bin Yin

Abstract

The availability of whole genomic sequences provides a great framework for biologists to address a broad range of scientific questions. However, functions of most mammalian genes remain obscure. The forward genetics strategy of insertional mutagenesis uses DNA mutagens such as retroviruses and transposable elements; this strategy represents a powerful approach to functional genomics. A variety of methods to uncover insertion sites have been described. This chapter details SplinkTA-PCR and SplinkBlunt-PCR, modified from splinkerette-PCR, for mapping chromosomally the insertion sites of a murine leukemia virus that causes leukemia in the BXH-2 strain of mice. These protocols are easy to use, reliable, and efficient.

Key words: Functional genomics, Insertion mutagenesis, Murine leukemia virus, PCR, Oncogene, Tumor suppressor gene, Leukemia

1. Introduction

With the whole genome sequences of human and major model organisms available, biological processes can be explored in an unprecedented way – questions with significantly increased broadness and complexity are pursued on a large scale and at a systemic level. On the wave of interest in dissection of gene functions and biological pathways, chromosome insertional strategies using DNA elements such as retroviruses and transposons represent a powerful forward genetics approach to functional genomics, among other induced or engineered mutagenesis (1–5). These DNA mutagens integrate into host genome DNA and can alter gene function. Through the identification of the chromosomal insertion sites of the mutagens, a gene can be assigned a certain

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_3, © Springer Science+Business Media, LLC 2011

phenotype-related function. Recently, insertional mutagenesis has been demonstrated to be fruitful in genome-wide analysis of cancer genes (6–13). The BXH-2 strain of mice constitutes a model of human acute myeloid leukemia (AML), which arises from infection by a murine leukemia virus (MuLV) (14). The BXH-2 MuLV not only acts as an insertional DNA mutagen to cause leukemia but also serves as a tag for leukemia-associated genes.

A variety of methods used to uncover insertion sites have previously been described, including genomic DNA library screening (15), ligation-mediated PCR (6), inverse PCR (16), VISA technique (17), T-linker PCR (18), and single nucleotide polymorphism-based mapping (19). These methods have been adopted to generate a large amount of insertion sites; however, their inherent limitations are posed either by excessive laborious work, low cloning efficiency, restriction site-related cloning bias, or nonspecific amplification. To facilitate functional genomic studies and maximize information that can be delivered from the use of insertional mutagenesis strategy, it is crucial to capture as complete a genome-wide profile of insertion sites as possible. The protocols that will be elaborated in this chapter provide a time-saving tool allowing for less laborious, less biased, and more efficient mapping of insertion sites.

1.1. Splinkerette-PCR

The splinkerette-PCR is derived from vectorette-PCR (20), also similar in principle to cassette ligation-anchored PCR (21), single-specific-primer PCR (22), rapid amplification of cDNA ends (23, 24), and rapid amplification of genomic DNA ends (25). This PCR begins with the digestion of genomic DNA with a restriction enzyme, followed by ligation of a double-stranded DNA linker, namely splinkerette, to the digested DNA. The insertion flanking sequences are then PCR amplified using a pair of primers specific to the integrated DNA and the splinkerette, respectively (26). Splinkerette-PCR features a hairpin structure present in the splinkerette, which helps overcome the undesired amplification, so-called end-repair priming phenomenon that decreases the specificity of conventional linker-mediated PCR. Another advantage of the approach is the elimination of the requirement for circularization that can often be a problematic step in some types of PCR. Splinkerette-PCR has been used to characterize genomic integration of provirus, transposon, and gene trap vector (6, 13, 27, 28). However, there is still a need for the development of more sophisticated and streamlined protocols that better accommodate the emerging large-scale insertion site mapping efforts. In line with this trend, a Web-based automated analysis and mapping of insertional mutagenesis sequence data has recently become available (29).

1.2. SplinkTA-PCR

In our experience of cloning retroviral insertion sites, splinkerette-PCR worked better than inverse PCR; however, splinkerette-PCR

mostly produced smears – there were rarely discernible discrete PCR bands for recovery and cloning. This is likely due to excessive nonspecific amplification, which is consistent with other investigators' observations (18). We also tried to clone the resulting smear PCR products for sequencing, only to find a very low percentage of positive colonies and an even lower number of specific PCR amplicons. With inverse-PCR, we observed even fewer PCR bands. In order to achieve saturation recovery of insertion sites, a more efficient approach is needed. SplinkTA-PCR (STA-PCR) is developed from splinkerette-PCR, and its strategy is illustrated in Fig. 1a (30). This PCR approach starts from digestion of genomic DNA with a cocktail of restriction enzymes generating 5′ overhang ends. The digested DNA is purified and modified at the 3′ end with addition of an adenosine, by taking advantage of the terminal nontemplated extension capacity of conventional *Taq* DNA polymerases (31, 32). This *A*ddition step is followed by ligation to SplinkTA, a modified splinkerette linker, which is made up of Splinkerette (the hairpin oligo) and PrimerLongTA (the oligo with an extra thymidine at 3′ end). Primary PCR is then performed on the ligation products as templates using primers specific to BXH-2 MuLV and SplinkTA, respectively. In secondary PCR, templates are switched to primary PCR products, and primers to a nested pair of primers specific to the long-terminal repeats of BXH-2 MuLV and complementary to PrimerLongTA, respectively. Individual secondary PCR bands are resolved on agarose gel and subsequently recovered for sequencing.

1.3. SplinkBlunt-PCR

SplinkBlunt-PCR is also modified from splinkerette-PCR (30). In this PCR method, as illustrated in Fig. 1b, genomic DNA is first digested with a cocktail of restriction enzymes. Next, a biotinylated primer is annealed to the digests to drive the primer extension reaction. *Pfu* DNA polymerase is used to produce blunt-end double-stranded DNA fragments. The extension products are subsequently purified with streptavidin magnetic particles, prior to ligation to SplinkBlunt, which is made by mixing Splinkerette with PrimerLong (the oligo without extra thymidine at the 3′ end), using the same procedure as in the STA-PCR protocol. Following ligation, the biotin-bearing DNA fragments are captured magnetically, denatured, and separated from other unbound DNA fragments. The supernatant is collected and used as templates for subsequent primary and secondary PCRs, which are performed similarly to what was described in STA-PCR Protocol. The downstream procedure following amplification is the same as that in STA-PCR.

STA-PCR and SplinkBlunt-PCR can consistently recover more insertion sites of BXH-2 MuLV with high specificity (for an example, see Fig. 2). The high efficiency of SplinkTA-PCR may be attributed to the integration of the following improvements: (1) Use of multiple enzymes rather than a single enzyme

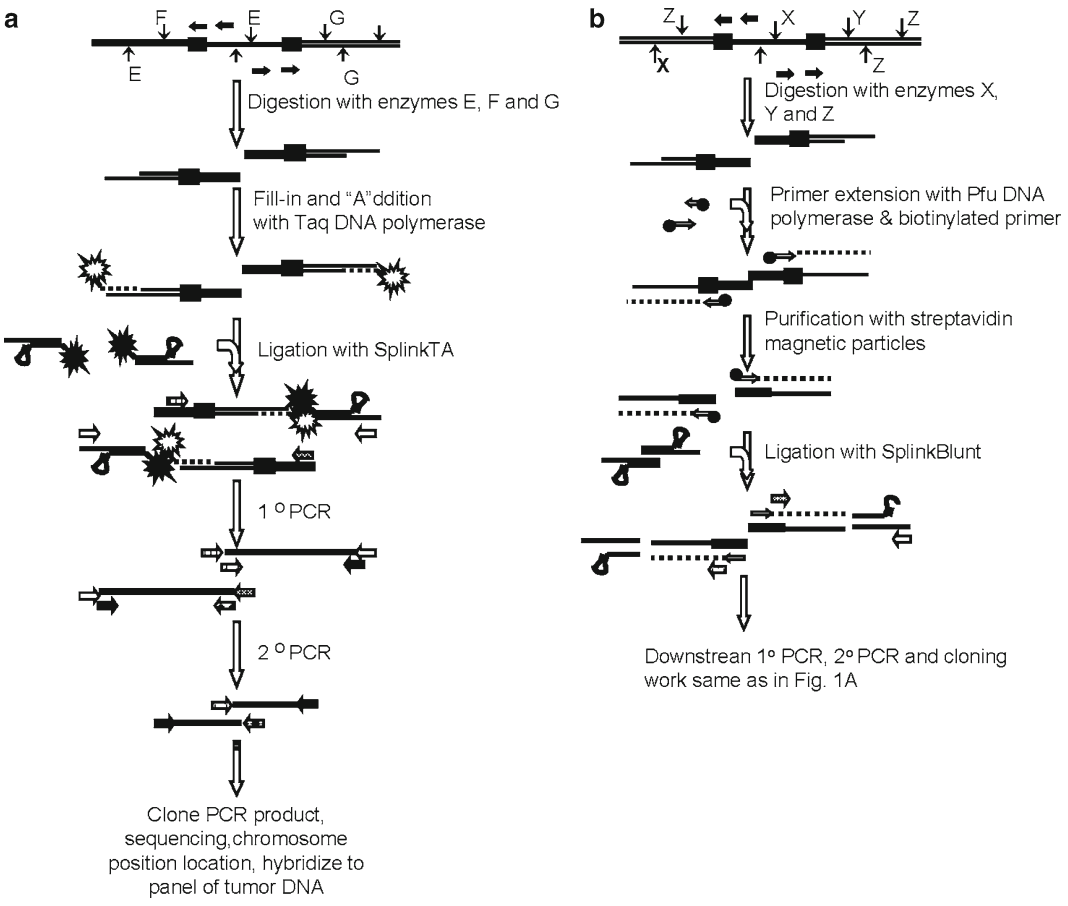


Fig. 1. Schematic of strategies for cloning genomic DNA sequence flanking a proviral insertion. (**a**) STA-PCR protocol. Genomic DNA (*horizontal lines*) is digested with a cocktail of restriction enzymes (noted as E, F, and G), which have no recognition site downstream to the primers (*horizontal arrow heads*) used for primary and secondary PCR within proviral genome sequence (*black boxes*). The digestion produces 5′ overhang ends that are subsequently filled in and added an adenosine at the ends with Taq DNA polymerase. The filled-in products are ligated to SplinkTA, a modified Splinkerette with a thymidine (*empty star*) at the end, which can exactly pair to the adenosine (*filled star*) at the end of extension products. Then two consecutive rounds of PCR are performed followed by PCR products cloning and sequencing. (**b**) SplinkBlunt-PCR protocol. Genomic DNA (*horizontal lines*) is digested with restriction enzymes (X, Y, and Z in this case) in a similar way described in (**a**). After the digested double-stranded DNA is denatured, proviral genome sequence-specific primer, which is biotinylated at its 5′ end (*filled circle*), is annealed to its cognate region on single-stranded DNA fragments and immediately extended with Pfu DNA polymerase to produce blunt ends that are subsequently ligated to SplinkBlunt. Those biotin-bearing DNA fragments are then purified with streptavidin magnetic particles prior to two rounds of PCR and products cloning as performed in (**a**) (Taken from Yin, B. et al., BioTechniques 2007. © 2008 BioTechniques. Used with permission).

for one digestion reaction gives rise to more readily amplifiable DNA fragments, reduces insertion cloning bias introduced by single restriction enzyme-dependent digestion, and saves time and effort by alleviating the requirements of more reactions and linkers which are necessary for single enzyme-based protocols; (2) Adoption of splinkerette eliminates the end-repair priming phenomenon and thereby results in greater efficiency

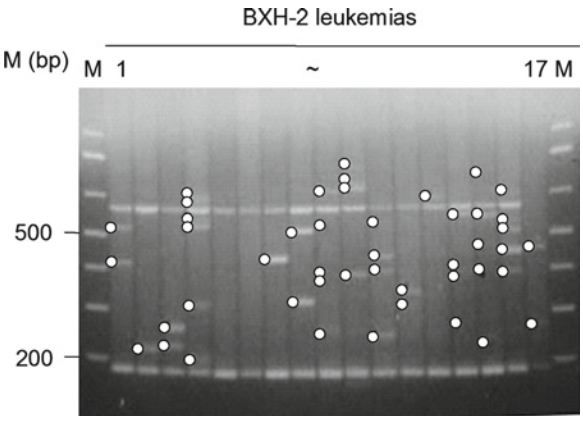


Fig. 2. Application of PCR-based protocols to large-scale screen of proviral insertion patterns in BXH-2 leukemias. *Lanes 1–17* are leukemia samples. The *white dots* denote the PCR products representing variant somatic acquired proviral insertions. M: 100-bp DNA ladder (Taken from Yin, B. et al., BioTechniques 2007. © 2008 BioTechniques. Used with permission).

in ligation-mediated PCR (26); (3) Formation of the SplinkTA in STA-PCR with an extended thymidine at the end increases ligation efficiency, maximizes its compatibility with a variety of digestion fragments generated by different restriction enzymes, and decreases the likelihood of forming tandem blunt ligation products; (4) Inclusion of primer extension and purification steps in SplinkBlunt-PCR eliminates the majority of nontemplate DNA that would adversely affect efficiency of downstream PCR, since the magnetic beads used in the purification only bind and retain double-stranded DNA fragments formed via streptavidin-labeled primers.

When comparing between STA-PCR and SplinkBlunt-PCR, the former has fewer steps and reduced likelihood of formation of tandem ligation. SplinkBlunt-PCR is characterized by cleaner PCR gel background and no limitation in choice of restriction enzymes to 5′ overhang-producing enzymes because both 3′ overhang and blunt DNA fragments can be primer extended. In addition, although comparable efficiency was observed for STA-PCR and SplinkBlunt-PCR in cloning of BXH-2 insertion, there are slight differences between these two methods in the recovery profile of insertions (see Fig. 3).

The ability of STA-PCR and SplinkBlunt-PCR to recover insertion sites could be enhanced by digesting genomic DNA with more than one combination of restriction enzymes. We have successfully tested various combinations of restriction enzymes, including: (1) *Aat* II, *Mfe* I, *Nde* I, (2) *Hind* III, *Pvu* II, *Xho* I, or (3) *Ase* I, *Bgl* I, *Eag* I, with different resultant PIS patterns. Further improvement of PIS recovery can be made by running through the cloning procedures for both upstream and downstream of the DNA integrated. We expect that with minor

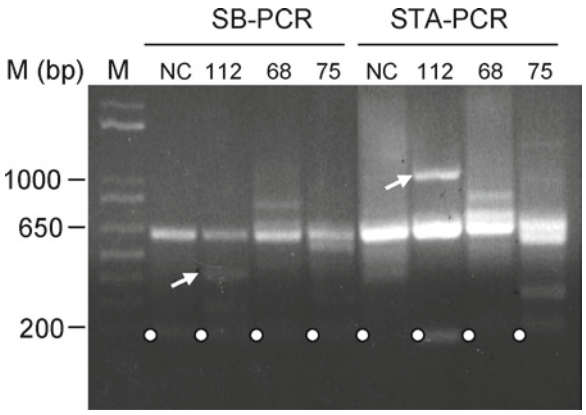


Fig. 3. Comparison of PCR band patterns amplified from leukemias using SplinkBlunt-PCR and STA-PCR protocols. B112, B68, and B75 are BXH-2-derived leukemia samples. NC: healthy BXH-2 mice tail DNA used as a control. M: 100-bp DNA marker. The *white arrows* indicate the differential bands. The size of DNA marker in base-pair is noted to the *left* (Taken from Yin, B. et al., BioTechniques 2007. © 2008 BioTechniques. Used with permission).

modification, these protocols can be readily adapted to identify insertion sites of other types of insertional mutagens, such as transposons, retrotransposons, etc. The techniques can also be applied to determine endpoints of genomic DNA fragments, chromosomal breakpoints involved in deletion or translocation, intron–exon junctions and gene regulatory regions.

2. Materials

2.1. Extraction of Genomic DNA from Cell Culture or Tissues

1. TE buffer: 10 mM Tris–HCl (pH 8.0) and 5 mM EDTA (pH 8.0). Stored at 4°C.
2. STE buffer: 1 mM EDTA, 10 mM Tris–HCl (pH 8.0) and 0.5 M NaCl. Stored at room temperature.
3. Proteinase K: dissolved in TE buffer at 10 mg/mL. Store aliquots at –20°C.
4. RNase A: dissolved in TE buffer at 2 mg/mL. Store aliquots at –20°C.
5. Lysis buffer: mix 4.6 mL of STE, 100 µL of 0.5 M EDTA, and 200 µL of 10% sodium dodecyl sulfate (SDS). Add 100 µL of 10 mg/mL proteinase K and 20 µL of 2 mg/mL RNase A prior to use (see Note 1).
6. Saturated phenol (pH ~7.9): Stored at 4°C. Avoid light exposure.
7. Chloroform: Stored in hazardous chemical carbine.
8. Phenol:chloroform (1:1): Stored at 4°C. Avoid light exposure.

9. Isopropanol: Stored in hazardous chemical carbine.
10. 70% Ethanol: Stored at 4°C.

2.2. STA-PCR (Including DNA Digestion, Extension, Ligation, PCR Amplification, and Examination)

1. Restriction enzyme and digestion buffer: *BspLU11* I, *Bcl* I, and *Taq* I (New England Biolabs). Stored at −20°C except for *BspLU11* I that is stored at −80°C (see Note 2).
2. Silicon oil.
3. QIAquick Nucleotide Purification Kit (Qiagen).
4. QIAquick PCR Purification Kit (Qiagen).
5. Splinkerette oligonucleotides (see Table 1 for sequences). Resuspended in TE buffer to the concentration of 100 μM. Store at −20°C.
6. T4 DNA ligase and ligation reaction buffer (Promega). Store at −20°C.
7. PCR grade water or Millipore purified water.
8. Biolase *Taq* DNA polymerase (Bioline USA Inc.). Store at −20°C.
9. 10× PCR reaction buffer and 50 mM $MgCl_2$, supplied with Biolase (Bioline USA Inc., Taunton, MA).
10. dNTPs: Resuspended in PCR grade water at 10 mM for each deoxynucleotide. Store aliquots at −20°C.
11. PCR primers. Resuspended in TE buffer at 100 μM and diluted further to 20 μM as working solutions (see Table 1). Store at −20°C.
12. Agarose powder.
13. Ethidium bromide at 10.0 mg/mL (Invitrogen).

Table 1
Sequence (5′→3′) of oligonucleotides used in STA-PCR and SplinkBlunt-PCR

1. Splinkerette	phos-CATGGTTGTTAGGACTGGAGGGGAAATCAATCCCCT
2. PrimerLongTA	CCTCCACTACGACTCACTGAAGGGCAAGCAGTCCTAACAACCATGT
3. PrimerLong	CCTCCACTACGACTCACTGAAGGGCAAGCAGTCCTAACAACCATG
4. Biotin-AKVp7711	Biotin-TCCAGGCTGCCATGCACGATGAC
5. AKVp7711	TCCAGGCTGCCATGCACGATGAC
6. AKVp7774	CTTTGACCTCCTTGTCCGAAGTA
7. AKVp8712	GCCAGTCCTCCGATAGACTGAG
8. AKVp875	CCAATAAAGCCTTTTGCTGTTGC
9. P-short	CCTCCACTACGACTCACTGAAGGGC
10. P-nested	GGGCAAGCAGTCCTAACAACCATG

14. 50× TAE stock solution: 0.242 g/L Tris Base, 5.7% glacial acetic acid, 50 mM EDTA (pH 8.0).
15. 100-bp DNA ladder (Invitrogen). Store at –20°C.
16. GeneAmp PCR system 9700 (PE Applied Biosystem).

2.3. SplinkBlunt-PCR (Including DNA Digestion, Extension, Ligation, PCR Amplification, and Examination)

1. Restriction enzyme and digestion buffer: *BspLU11* I, *Bcl* I and *Taq* I, or other enzymes of your choice (New England Biolabs). Stored at –20°C except for *BspLU11* I that is stored at –80°C.
2. Silicon oil.
3. QIAquick PCR Purification Kit (Qiagen).
4. *Pfu* DNA polymerase (5 U/μL) and 10× PCR Buffer (Stratagene).
5. dNTPs: Prepared and stored as described in Subheading 2.2.
6. Biotinylated primer Bio-AKVp7711 (see Table 1 for sequences).
7. Microcon YM-30 column (Millipore).
8. Streptavidin-coupled Magnetic Particles (Roche).
9. Magnetic Particle Concentrator (MPC) (Roche).
10. 2× BW Buffer: containing 10 mM Tris–HCl (pH 7.5), 1 mM EDTA, and 2 M NaCl.
11. 0.1 N NaOH.

Other reagents required are the same as listed in Subheading 2.2.

2.4. PCR Product Cloning and Sequencing

1. In-gel DNA purification kit (Qiagen) (see Note 3).
2. TOPO-TA cloning vector kit. Store at –20°C (see Note 4).
3. TOP10 competent bacteria (Invitrogen). Store at –80°C (see Note 5).
4. Luria Bertani (LB) medium: 10 g/L of tryptone, 5 g/L of yeast extract, and 10 g/L of NaCl (Note: For bacteria culture) (see Note 6).
5. S-gal LB powder (Sigma, St. Louis, MO) (see Note 7).
6. Solution I: 50 mM glucose, 25 mM Tris–HCl (pH 8.0), and 10 mM EDTA (pH 8.0). Filter sterilized. Store at 4°C.
7. Solution II: 0.1 M NaOH and 1% SDS. Store at 4°C.
8. Solution III: 3 M potassium acetate and 5 M glacial acetic acid. Store at 4°C.

3. Methods

3.1. Extraction of Genomic DNA from Cell Culture or Tissues

1. Collect tissues or pellet cells into 15-mL conical polypropylene tubes (see Note 1). Add appropriate amount of lysis buffer for DNA extraction and incubate at 56°C overnight with shaking.

2. Add an equal amount of Tris-buffered phenol to samples. Mix thoroughly by inversion. Centrifuge at 2,500×*g* for 5–10 min.
3. Transfer aqueous phase (top layer) to new 15-mL conical polypropylene tubes using pipet tips cut at the end to avoid genomic DNA shearing. Take care not to take any undissolved tissue or white precipitates between layers.
4. Add equal volume of phenol:chloroform (1:1) and mix thoroughly by inversion for 5–10 min.
5. Centrifuge at 2,500×*g* for 5–10 min and transfer the aqueous phase to new 15-mL conical polypropylene tubes.
6. Add equal volume of chloroform and mix thoroughly by inversion for 5–10 min.
7. Centrifuge at 2,500×*g* for 5–10 min and transfer the aqueous phase to new 15-mL conical polypropylene tubes. If the aqueous phase is not clear, repeat step 2.
8. Add 0.7 volume of isopropanol and mix thoroughly by gentle inversion.
9. Genomic DNA is usually visible at this step as fluffy or silky white material. Proceed to transfer the DNA into a new 1.5-mL centrifuge tube with a pipet tip and allow to air dry. If DNA pellet is not visible, centrifuge at 2,500×*g* for 10 min at 4°C to pellet genomic DNA.
10. Discard supernatant and wash with 70% ethanol.
11. Centrifuge at 2,500×*g* for 10 min at 4°C to pellet genomic DNA again.
12. Discard supernatant and add 100–300 µL of TE buffer.
13. Incubate at 56°C for 30 min to let genomic DNA dissolve into the TE buffer. For a more complete dissolve, incubate at 37°C overnight.
14. Determine genomic DNA concentration using a spectrophotometer. DNA can then be aliquoted and stored for long term at –20°C.

3.2. STA-PCR (Including DNA Digestion, Addition, Ligation, PCR Amplification, and Examination) (see Notes 8–10)

1. Digestion: Digest 2 µg of genomic DNA with 7.5 U of *Taq* I at 65°C for 4 h in a final volume of 50 µL, followed by another 4-h digestion with 7.5 U *BspLU11* I and 10 U *Bcl* I at 48°C (see Notes 11–13).
2. Purification: Purify the digestion products (from step 1) with QIAquick PCR Purification Kit. According to the directions provided by the kit, add 250 µL Buffer PB to each sample and apply the mixture to the mini columns. Centrifuge at 11,000×*g* for 30 s. Wash with 750-µL Buffer PE at 11,000×*g* for 30 s and centrifuge the column for an additional 60 s at maximum speed. Add 30-µL H_2O to the columns and spin to elute the DNA bound to the membrane.

3. *A*ddition: To each purified digestion (from step 2), add the following:

H_2O	12.5 µL
10× PCR buffer (without Mg^{2+})	5.0 µL
50 mM $MgCl_2$	1.5 µL
10 mM dNTPs	1.0 µL
5 U/µL *Taq* DNA polymerase	0.5 µL

Incubate at 72°C for 20 min

(see Notes 14 and 15).

4. Purification: purify the extension reaction with QIAquick PCR Purification Kit as described in step 2.
5. Making SplinkTA: mix 20 µL of 100 µM Splinkerette and 20 µL of 100 µM PrimeretteLongTA. Incubate at 80°C for 5 min. Cool to room temperature (20–25°C) (see Note 16).
6. Ligation: To 5 µL of each purified sample (from step 4), add the following:

H_2O	1.75 µL
10× Ligation buffer	1.0 µL
50 pmoL/µL SplinkerTA	2.0 µL
200 U/µL T_4 DNA ligase	0.25 µL

Incubate at 16°C overnight.

Note. For multiple samples, prepare a master mix.

7. Purification: purify the ligation reaction with QIAquick PCR Purification Kit as described in step 2.
8. Primary PCR: In 0.2 mL thin-wall PCR tubes, compose the following components for one reaction:

PCR grade water	37.0 µL
10× PCR buffer	5.0 µL
50 mM $MgCl_2$	1.5 µL
10 mM dNTPs	1.0 µL
20 µM primer AKVp7711	1.0 µL
20 µM primer P-short	1.0 µL
5 U/µL *Taq* DNA polymerase	0.5 µL
Purified ligation products as templates (from step 7)	3.0 µL

The primary PCR cycling conditions are as follows:

(a) 95°C for 1 min 30 s

(b) 10 cycles of
- 95°C for 5 s
- 70°C for 3 min 10 s with decreasing 0.5°C each cycle

(c) 20 more cycles of
- 95°C for 5 s
- 65°C for 3 min 10 s

(d) Final extension at 70°C for 10 min

(e) 4°C hold

9. Secondary PCR: all components in each reaction are the same as those in primary PCR (step 8) except for using primary PCR products as templates and replacing the pair of primers with nested primers AKVp8712 and P-nest (see Note 17).

PCR grade water	39.0 µL
10× PCR buffer	5.0 µL
50 mM $MgCl_2$	1.5 µL
10 mM dNTPs	1.0 µL
20 µM primer AKVp8712	1.0 µL
20 µM primer P-nest	1.0 µL
5 U/µL *Taq* DNA polymerase	0.5 µL
Primary PCR products as templates (from step 8)	1.0 µL

The amplification parameters in secondary PCR are changed to:

(a) 95°C for 1 min

(b) 11 cycles of
- 94°C for 15 s
- 70°C for 2 min 40 s with decreasing 0.6°C each cycle

(c) 25 more cycles of
- 94°C for 15 s
- 64°C for 30 s
- 70°C for 2 min 10 s

(d) Final extension at 70°C for 10 min

(e) 4°C hold

10. Check secondary PCR results by resolving 10 µL of each sample in 1% agarose gel and examining on a UV-light box.

3.3. SplinkBlunt-PCR (Including DNA Digestion, Extension, Ligation, PCR Amplification, and Examination) (See Notes 9 and 18)

1. Digestion: 2 µg of genomic DNA are incubated with *Taq* I at 65°C for 4 h followed by another 4-h digestion with *BspLU11* I and *Bcl* I at 48°C (see Notes 11, 13, 19, and 20).
2. Purification: Purify the extension reaction with QIAquick PCR Purification Kit as described earlier in STA-PCR (step 2).

3. Extension: To each purified digestion (from step 2), add the following:

H_2O	39.5 µL
10× *Pfu* PCR buffer (with Mg^{2+})	5.0 µL
10 mM dNTPs	1.0 µL
5 pmol/µL Bio-AKVp7711 (see Table 1 for sequence)	4.0 µL
2.5 U/µL *Pfu* DNA polymerase	0.5 µL

Incubate at 95°C for 5 min, 64°C for 30 min, and 72°C for 20 min.

Note. For multiple reactions, making mastermix is highly recommended.

4. Purification: Purify the extension reaction with MicroconYM-30 column according to the manufacturer's instructions. Briefly, add 450 µL of H_2O to each sample, and transfer to the column. Centrifuge at 14,000 × *g* for 10 min. Place the column upside down and centrifuge for 3 min.
5. DNA capture: To recover the Bio-AKVp7711-derived extension products, wash 10 µL of Streptavidin-coupled Magnetic Particles per sample with 50 µL of 2× BW Buffer. Resuspend the particles in 40 µL of 2× BW Buffer, then mix with the purified extension products (from step 4), and incubate at room temperature (20–25°C) for 1–3 h. Capture the biotin-bearing DNA fragments using Magnetic Particle Concentrator, and wash twice with H_2O.
6. Making SplinkerBlunt: Mix 20 µL of 100 µM Splinkerette and 20 µL of 100 µM PrimeretteLong. Incubate at 80°C for 5 min. Cool down to room temperature (20–25°C) (see Note 16).
7. Ligation: The captured DNA was ligated to the SplinkerBlunt. To 5 µL of each purified sample (from step 5), add the following:

H_2O	1.75 µL
10× Ligation buffer	1.0 µL
50 pmoL/µL SplinkerBlunt	2.0 µL
200 U/µL T_4 DNA ligase	0.25 µL

Incubate at 16°C overnight.

Note. For multiple samples, prepare a master mix.

8. Recapturing DNA: After ligation, the biotin-bearing DNA fragments (from step 7) are captured again in MPC and washed with 100 µL of H_2O. Denature the recaptured DNA in 5 µL of 0.1 N NaOH at room temperature for 10 min. Separate the magnetic particles from released DNA fragments

with MPC. Save the supernatant as templates for subsequent primary and secondary PCR.

9. Primary PCR: In 0.2 mL thin-wall PCR tubes, compose the following components for one reaction:

PCR grade water	37.0 μL
10× PCR buffer	5.0 μL
50 mM $MgCl_2$	1.5 μL
10 mM dNTPs	1.0 μL
20 μM primer AKVp7774	1.0 μL
20 μM primer P-short	1.0 μL
5 U/μL *Taq* DNA polymerase	0.5 μL
Purified ligation products as templates (from step 8)	3.0 μL

The primary PCR cycling conditions are as follows:

(a) 95°C for 1 min 30 s

(b) 10 cycles of
- 94°C for 15 s
- 70°C for 3 min 10 s with decreasing 0.6°C each cycle

(c) 20 more cycles of
- 94°C for 15 s
- 64°C for 30 s
- 70°C for 2 min 30 s

(d) Final extension at 70°C for 10 min

(e) 4°C hold

10. Secondary PCR: Performed as described in STA-PCR Protocol (step 9).
11. Check secondary PCR results by resolving 10 μL of each sample in 1% agarose gel and examining on a UV-light box.
12. An example of large-scale cloning PCR results produced is shown in Fig. 2 (see Note 21).
13. A comparison of STA-PCR and SplinkBlunt-PCR results is shown in Fig. 3 (see Note 22).
14. A comparison of SplinkBlunt-PCR and Southern blotting result is shown in Fig. 4 (see Note 23).

3.4. PCR Product Cloning and Sequencing

1. Run secondary PCR products on preparative 1% agarose gel containing ethidium bromide (see Note 24).
2. Cut out bands with sterile blades and recover DNA using QIAquick Gel Purification kit following the manufacturer's instructions.
3. Use TOPO TA Cloning Kits according to the manual exactly. Ligate 4 μL of recovered PCR bands with 1 μL of TOPO-TA

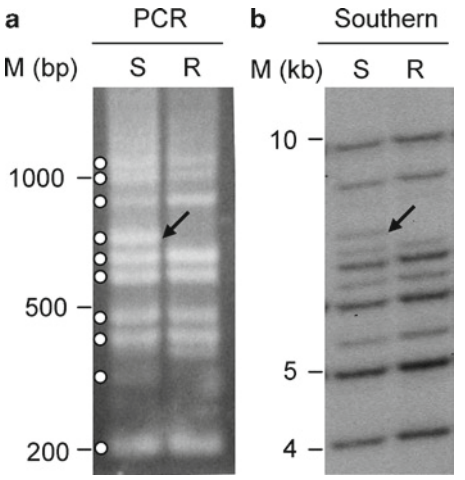


Fig. 4. Comparison between SplinkBlunt-PCR and Southern blot assay in detecting differential proviral insertions. The genomic DNA from Ara-C-sensitive cell line B117P (*lane S*) as a passage control, or Ara-C-resistant cell line B117H (*lane R*), was subjected to proviral insertion cloning PCR protocol (**a**) or Southern blotting analysis (**b**). *Arrows* indicate the differential insertion (Taken from Yin, B. et al., BioTechniques 2007. © 2008 BioTechniques. Used with permission).

cloning vector in the presence of 1 μL of 6× ligation reaction buffer, transform the TOP10 competent cells, and spread bacteria culture onto S-gal LB plates containing the appropriate antibiotic for the cloning vector used. For each PCR band, 3–5 white colonies are typically picked to determine positive transformants (see Notes 25–27).

4. Plasmid DNA is prepared from positive colonies for sequencing using primer AKVp8753.
5. DNA sequences are aligned with Ensembl Mouse Genome Database to map the chromosomal position of insertion sites.

4. Notes

1. Recipe is given here for $2–10 \times 10^7$ cells or 100–200-mg tissues; it can be scaled up or down for more or less cells/tissues.
2. Any restriction enzymes that produce 5′ overhang should, in principle, work equally well. Indeed, we tried other enzymes generating 5′ overhang and obtained different PCR band patterns. We therefore recommend comparing various combinations of enzymes before deciding which enzymes to use in a particular insertion site cloning work.
3. Other economic kits, such as glass binding-based methods, also work very well. We found that the Qiagen kit designed for purification of DNA from agarose gel is relatively convenient, giving rise to high yield and clean PCR products.

4. For cloning of large PCR products, TOPO-TA-XL works better for us.
5. For transformation of ligated insert/vector DNA: In our hands, the TOP10 competent cells from Invitrogen give rise to efficient transformation; however, they are relatively expensive. To reduce the cost, the use of smaller than recommended amount of cells, as low as 25 μL per transformation, works equally well.
6. Use of the S.O.C. medium is optional here, which contains 2% tryptone, 0.5% yeast extract, 2.5 mM KCl, 10 mM NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose with pH adjusted to 7.0 ± 0.2. The S.O.C. bacteria culture medium generally works better, producing more colonies after transformation.
7. Making LB plates: Add the appropriate antibiotics for selection for transformed colonies.
8. The STA-PCR procedure typically requires less than 7 days for completion of mapping insertions.
9. Take great care to prevent PCR cross-contamination. Common solutions include using aerosol-resistant filter pipet tips, dedicated PCR working area and reagents, and their UV-light exposure, etc.
10. When compared with the T-linker PCR, STA-PCR features the combined advantages of adoption of splinkerette, use of an enzyme cocktail, and fewer experimental steps. In addition, since STA-PCR involves complete digestion with 5′ overhang enzymes, there is no limitation by the need to perform the incomplete digestion – which can often go out of control – with 3′ overhang enzymes required by T-linker PCR.
11. Digestion with multiple restriction enzymes (*Taq* I, *Bcl* I, and *BspLU11* I) not only gives rise to shorter DNA fragments on average than digestion with a single enzyme and renders the fragments more readily PCR amplified, but also reduces cloning bias introduced by single restriction enzyme-based methods as were used in previous reports. This is an important feature that saves time and effort by alleviating the need for extra reactions and types of linkers necessary for single enzyme approaches. This is also likely to increase the specificity of PCR.
12. Choice of restriction enzyme cocktails can be tailored to the particular interest of individual studies.
13. Use of silicon oil to cover the top of digestion solution is recommended to prevent reaction loss arising from liquid evaporation.
14. For multiple samples, making master mix is highly recommended for easier and consistent handling.
15. *Taq* DNA polymerases are used here in order to form sticky double-stranded DNA ends with an extruding adenosine through the terminal nontemplated adenosine addition reaction. To test whether this approach can increase ligation

efficiency, the *Taq*-treated digests were ligated to the pCR2.1-TOPO vector (Invitrogen) that has a 3′ thymidine, followed by transformation of competent cells. We found that the *Taq*-treated digests produced a higher number of transformants than did the untreated digests as a control.

16. To do this, we use a heating block preheated to 80°C, followed by turning it off and allowing it to slowly cool down to room temperature. Alternatively, this can be done using a PCR amplifier with a decreasing 0.5°C per min over 2 h.
17. If no discrete bands are visible on gel, we recommend trying to amplify 1:20–200 dilutions of primary PCR products in secondary PCR.
18. The procedure described here typically requires less than 10 days for completion of mapping insertions.
19. In principle, any other combination of restriction enzymes can be applied which do not cut the integrant DNA downstream to the primer used in the extension reaction (also see the description of extension reaction).
20. For a similar reason as with digestion in STA-PCR, it is recommended that comparison of different combinations of enzymes be made to select a favorable enzyme cocktail, prior to a large-scale insertion site cloning project.
21. The presence of two germline proviral insertions in BXH-2 tumors poses a heavy interference with isolation of somatic acquired proviral insertions, which made the proviral insertion cloning very challenging.
22. Because STA-PCR and SplinkBlunt-PCR give rise to different PIS patterns in some samples (for an example, see Fig. 3), use of both procedures would improve PIS recovery rate. In our experience, nonoverlapping PISs account for approximately 10% of total PISs isolated. Therefore, combination of the two protocols is highly recommended in some cases where a full or nearly full recovery is desired.
23. When comparing our PCR methods with Southern blotting analysis, we found a close-to-full recovery of insertions by PCR, with 197 insertions amplified out of 80 tumors, versus 68 insertions detected by Southern from 29 BXH2 tumors. Given the proviral insertion recovery rate achieved, these PCR protocols represent an improved method for proviral insertion site cloning in BXH-2 tumors.
24. To increase gel resolution for smaller PCR bands, up to 2% agarose gel can be readily used here.
25. Alternatively, sequence directly the DNA samples that are recovered from strong and pure bands with internal sequencing primer AKVp8753.

26. If the PCR bands are too faint, it is sometimes helpful to PCR reamplify the recovered DNA prior to TA cloning of the PCR products or direct sequencing.
27. The use of pCR®4-TOPO vector (Invitrogen) may significantly reduce transformation background because self-ligated vectors are lethal to competent cells.

Acknowledgments

The author would like to express thanks to Ms. Marianna L. Wong and Mr. John W. Myers III for their help with grammatical editing. This work is supported by the National Basic Research Program of China (973 Program, No. 2011CB933501, to Dr. Bin Yin) and the NSFC Project (No. 81070417, to Dr. Bin Yin).

References

1. Zambrowicz, B. P., Friedrich, G. A., Buxton, E. C., Lilleberg, S. L., Person, C., and Sands, A. T. (1998) Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392,** 608–11.
2. Friedrich, G. and Soriano, P. (1991) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* **5,** 1513–23.
3. Miskey, C., Izsvak, Z., Kawakami, K. and Ivics, Z. (2005) DNA transposons in vertebrate functional genomics. *Cell Mol Life Sci* **62,** 629–41.
4. Uren, A. G., Kool, J., Berns, A. and van Lohuizen, M. (2005) Retroviral insertional mutagenesis: past, present and future. *Oncogene* **24,** 7656–72.
5. Neil, J. C. and Cameron, E. R. (2002) Retroviral insertion sites and cancer: fountain of all knowledge? *Cancer Cell* **2,** 253–5.
6. Mikkers, H., Allen, J., Knipscheer, P., Romeijn, L., Hart, A., Vink, E., and Berns, A. (2002) High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. *Nat Genet* **32,** 153–9.
7. Suzuki, T., Minehata, K., Akagi, K., Jenkins, N. A. and Copeland, N. G. (2006) Tumor suppressor gene identification using retroviral insertional mutagenesis in Blm-deficient mice. *Embo J* **25,** 3422–31.
8. Suzuki, T., Shen, H., Akagi, K., Morse, H. C., Malley, J. D., Naiman, D. Q., Jenkins, N. A., and Copeland, N. G. (2002) New genes involved in cancer identified by retroviral tagging. *Nat Genet* **32,** 166–74.
9. Iwasaki, M., Kuwata, T., Yamazaki, Y., Jenkins, N. A., Copeland, N. G., Osato, M., Ito, Y., Kroon, E., Sauvageau, G., and Nakamura, T. (2005) Identification of cooperative genes for NUP98-HOXA9 in myeloid leukemogenesis using a mouse model. *Blood* **105,** 784–93.
10. Castilla, L. H., Perrat, P., Martinez, N. J., Landrette, S. F., Keys, R., Oikemus, S., Flanegan, J., Heilman, S., Garrett, L., Dutra, A., Anderson, S., Pihan, G. A., Wolff, L., and Liu, P. P. (2004) Identification of genes that synergize with Cbfb-MYH11 in the pathogenesis of acute myeloid leukemia. *Proc Natl Acad Sci U S A* **101,** 4924–9.
11. Lund, A. H., Turner, G., Trubetskoy, A., Verhoeven, E., Wientjens, E., Hulsman, D., Russell, R., DePinho, R. A., Lenz, J., and van Lohuizen, M. (2002) Genome-wide retroviral insertional tagging of genes involved in cancer in Cdkn2a-deficient mice. *Nat Genet* **32,** 160–5.
12. Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., and Xu, T. (2005) Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122,** 473–83.
13. Starr, T. K., Allaei, R., Silverstein, K. A., Staggs, R. A., Sarver, A. L., Bergemann, T. L., Gupta, M., O'Sullivan, M. G., Matise, I., Dupuy, A. J., Collier, L. S., Powers, S., Oberg, A. L., Asmann, Y. W., Thibodeau, S. N., Tessarollo, L., Copeland, N. G., Jenkins, N. A., Cormier, R. T., and Largaespada, D. A. (2009) A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science* **323,** 1747–50.

14. Jenkins, N. A., Copeland, N. G., Taylor, B. A., Bedigian, H. G., and Lee, B. K. (1982) Ecotropic murine leukemia virus DNA content of normal and lymphomatous tissues of BXH-2 recombinant inbred mice. *J Virol* **42,** 379–88.
15. Blaydes, S. M., Kogan, S. C., Truong, B. T., Gilbert, D. J., Jenkins, N. A., Copeland, N. G., Largaespada, D. A., and Brannan, C. I. (2001) Retroviral integration at the Epi1 locus cooperates with Nf1 gene loss in the progression to acute myeloid leukemia. *J Virol* **75,** 9427–34.
16. Triglia, T., Peterson, M. G., and Kemp, D. J. (1988) A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res* **16,** 8186.
17. Hansen, G. M., Skapura, D., and Justice, M. J. (2000) Genetic profile of insertion mutations in mouse leukemias and lymphomas. *Genome Res* **10,** 237–43.
18. Yan, Y., Li, L., Gu, J., Tan, G., and Chen, Z. (2003) T-linker-specific ligation PCR (T-linker PCR): an advanced PCR technique for chromosome walking or for isolation of tagged DNA ends. *Nucleic Acids Res* **31,** e68.
19. Shen, H., Suzuki, T., Munroe, D. J., Stewart, C., Rasmussen, L., Gilbert, D. J., Jenkins, N. A., and Copeland, N. G. (2003) Common sites of retroviral integration in mouse hematopoietic tumors identified by high-throughput, single nucleotide polymorphism-based mapping and bacterial artificial chromosome hybridization. *J Virol* **77,** 1584–8.
20. Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J. C., and Markham, A. F. (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res* **18,** 2887–90.
21. Mueller, P. R. and Wold, B. (1989) In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246,** 780–6.
22. Shyamala, V. and Ames, G. F. (1989) Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR. *Gene* **84,** 1–8.
23. Frohman, M. A. (1993) Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. *Methods Enzymol* **218,** 340–56.
24. Schaefer, B. C. (1995) Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. *Anal Biochem* **227,** 255–73.
25. Mizobuchi, M. and Frohman, L. A. (1993) Rapid amplification of genomic DNA ends. *Biotechniques* **15,** 214–6.
26. Devon, R. S., Porteous, D. J. and Brookes, A. J. (1995) Splinkerettes – improved vectorettes for greater efficiency in PCR walking. *Nucleic Acids Res* **23,** 1644–5.
27. Yin, B., Delwel, R., Valk, P. J., Wallace, M. R., Loh, M. L., Shannon, K. M., and Largaespada, D. A. (2009) A retroviral mutagenesis screen reveals strong cooperation between Bcl11a overexpression and loss of the Nf1 tumor suppressor gene. *Blood* **113,** 1075–85.
28. Horn, C., Hansen, J., Schnutgen, F., Seisenberger, C., Floss, T., Irgang, M., De-Zolt, S., Wurst, W., von Melchner, H., and Noppinger, P. R. (2007) Splinkerette PCR for more efficient characterization of gene trap events. *Nat Genet* **39,** 933–4.
29. Kong, J., Zhu, F., Stalker, J., and Adams, D. J. (2008) iMapper: a web application for the automated analysis and mapping of insertional mutagenesis sequence data against Ensembl genomes. *Bioinformatics* **24,** 2923–5.
30. Yin, B., and Largaespada, D. A. (2007) PCR-based procedures to isolate insertion sites of DNA elements. *Biotechniques* **43,** 79–84.
31. Clark, J. M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res* **16,** 9677–86.
32. Hu, G. (1993) DNA polymerase-catalyzed addition of nontemplated extra nucleotides to the 3′ end of a DNA fragment. *DNA Cell Biol* **12,** 763–70.

Chapter 4

Lariat-Dependent Nested PCR for Flanking Sequence Determination

Daniel J. Park

Abstract

Methods detailed in this chapter relate to the use of Lariat-dependent Nested (LaNe) PCR to characterize unknown RNA or DNA sequence flanking known regions. A multitude of approaches designed to determine flanking sequences have been described in the literature. Variously, problems related to these approaches include lack of resolution or failure, depending on experimental context, and complex handling. LaNe-based methods are designed to harness "two-sided" gene-specific PCR with the option of nesting but without the requirement for inefficient and involved enzyme preprocessing steps.

Key words: Flanking sequence determination, LaNe, PCR walking

1. Introduction

While relatively recent higher throughput DNA sequencing is revolutionizing the way biological research is conducted, there remain applications for more focused analyses of flanking sequence. The size of the region to be resolved, number of specimens, complexity and duration of processing and bioinformatic analysis, intractability of sequence to methodology, and cost are among the factors influencing the choice of approach to the study of flanking sequence. Many methods for PCR-based flanking sequence determination have been described, such as 3′Rapid Amplification of cDNA Ends (3′RACE), 5′RACE, Rapid Amplification of Genomic DNA Ends, Inverse PCR, CapSelect, Ligase-mediated PCR, Splinkerette PCR, Thermal Asymmetric Interlaced (TAIL) PCR, and Universal Fast Walking (1–11). Tonooka and Fujishima have recently reviewed PCR methods for

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_4, © Springer Science+Business Media, LLC 2011

walking along genomic DNA (12). Methods for PCR-based flanking sequence determination exhibit various limitations, including a requirement for relatively large quantities of input material, complex processing steps relying on multiple handling steps, and dependence on efficient enzyme steps involving restriction endonucleases, ligases, and/or exonuclease processing. In most cases, there is a reliance on the so-called "one-sided" PCR in which extension is primed by a gene-specific primer and a "general" primer targeted to sites common across the genome or adapters present in most species of the template mix. These approaches work satisfactorily in many cases, but some can encounter difficulties in more challenging contexts. In particular, one-sided approaches often yield "smeared products" as visualized following gel electrophoresis such that desired products of interest are difficult to resolve.

The Lariat-dependent Nested (LaNe)-based methods 3′RACE LaNe, 5′RACE LaNe, and LaNe RAGE (13–15) are derived from the "panhandle" concept employed in Universal Fast Walking (UFW) in which a lariat structure is involved in self-priming to yield a template which enables "two-sided" and nested PCR to be applied to flanking sequence determination. LaNe-based methods streamline the UFW concept workflow and system by removing the requirement for multiple rounds of exonuclease addition and "end-fill" steps prior to initiating "PCR-proper." Instead, LaNe-based self-priming is facilitated by denaturation and annealing steps inherent to conventional PCR thermocycling (the reaction scheme for LaNe RAGE is depicted in Fig. 1).

In this chapter, LaNe-based protocols are presented for 5′ or 3′ oriented RNA flanking sequence determination (5′RACE LaNe and 3′RACE LaNe) and genomic DNA flanking sequence determination (LaNe RAGE). Further, special requirements and tips for customization of the approaches are presented in Subheading 4.

2. Materials

2.1. 3′RACE LaNe

1. 20 ng total RNA (see Note 1).
2. 10 mM dNTPs.
3. ThermoScript™ Reverse Transcriptase and accompanying buffers (Invitrogen) (see Note 2).
4. TTT3 primer: 5′-gene-specific sequence followed by TTTTTTTTTTTTTTTTV-3′. The 5-prime gene-specific region of this primer should be in the same orientation as GSP1 and GSP4 (see Subheading 2.1, item 7) in alignment with position "3" as depicted in Fig. 1.
5. RNaseH (Invitrogen).

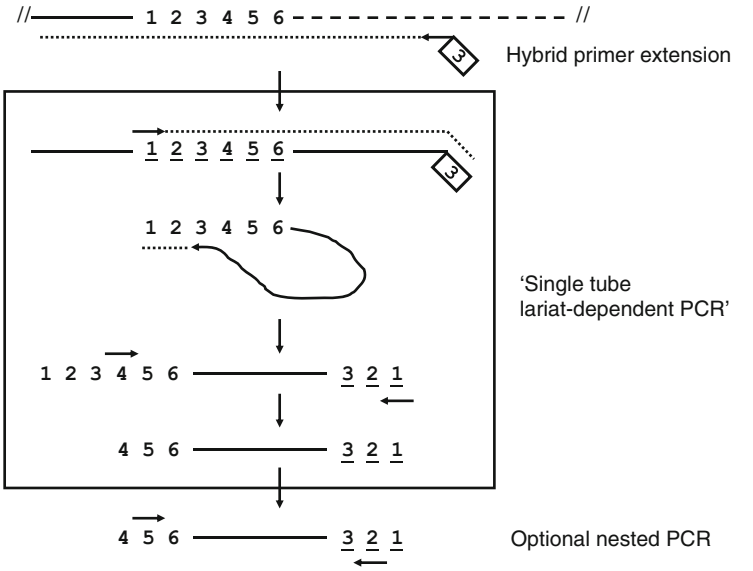


Fig. 1. Schematic illustration of the LaNe approach applied in LaNe RAGE. A hybrid primer containing gene-specific sequence at the 5′ end and nongene-specific sequence at the 3′ end is used to prime extension from the template (hybrid primer extension). The resulting product is then amenable to gene-specific second strand synthesis, followed by lariat structure formation and self-primed extension, to yield template which can be amplified by "two-sided PCR" using two gene-specific primers (Single tube lariat-dependent PCR). This can occur in a standard PCR thermocycling profile due to the multiple rounds of denaturation and annealing in the presence of the appropriate primers and thermostable DNA polymerase. The product of Single-tube lariat-dependent PCR can be applied as template in a nested PCR scheme (optional nested PCR). Numbers indicate features of a known sequence region. It is important to note that features 1–6 should always be oriented with feature 6 closest to the unknown sequence to be determined. The *dashed line* indicates the unknown sequence of the template which is to be determined. *Underlined numbers* indicate reverse complement sequences with respect to these features. Primers are represented by *horizontal arrows*. *Dotted lines* indicate polymerase extension.

6. *PfuTurbo*® DNA Polymerase and accompanying buffers and solutions (Stratagene) (see Note 3).
7. GSP1 and GSP4: Gene-specific primers 1 and 4 should be designed according to standard PCR primer conventions with respect to avoiding primer dimers and intraprimer annealing, attempting to prevent nonspecific priming, and ensuring that the T_m of the primers is a few degrees higher than the intended annealing temperature during thermal cycling *but* that both are in the same orientation as each other with the 3′ ends pointing towards the unknown region to be elucidated (refer to Fig. 1) (see Notes 4 and 5). GSP1 and GSP4 should be designed in alignment with positions "1" and "4," respectively, in Fig. 1.
8. GSP2 and GSP5: Gene-specific primers (optional – see Notes 6 and 7) should be designed in a similar manner to GSP1 and GSP4 but aligned with positions "2" and "5," respectively, in Fig. 1.

9. QIAEXII DNA purification system (Qiagen) (optional – see Note 8).
10. GSP6: Gene-specific primer (optional – see Note 8) should be designed in a similar manner to GSP1, GSP2, GSP4, and GSP5 but aligned with position "6" in Fig. 1.

2.2. 5′ RACE LaNe

2.2.1. First-Strand Intrastrand-Annealing 5′ RACE LaNe (FI RACE LaNe)

1. 200 ng total RNA (see Note 1).
2. 10 mM dNTPs.
3. ThermoScript™ Reverse Transcriptase and accompanying buffers (Invitrogen) (see Note 2).
4. RT primer: Reverse transcription primer should be oriented such that the 3-prime end is directed toward the unknown 5-prime mRNA region to be elucidated and reverse transcribed product includes features "1" to "6" of Fig. 1.
5. RNaseH (Invitrogen).
6. *PfuTurbo*® DNA Polymerase and accompanying buffers and solutions (Stratagene) (see Note 3).
7. GSP1 and GSP4: Gene-specific primers 1 and 4 (see Subheading 2.1, item 7 and Fig. 1). In this case, 3-prime ends of primers are all oriented towards the unknown 5-prime region of mRNA to be elucidated.
8. GSP2 and GSP5: Gene-specific primers 2 and 5 (see Subheading 2.1, item 8 and Fig. 1). In this case, 3-prime ends of primers are all oriented towards the unknown 5-prime region of mRNA to be elucidated.
9. QIAEXII DNA purification system (Qiagen) (optional – see Note 8).
10. GSP6: Gene-specific primer 6 (see Subheading 2.1, item 10 and Fig. 1) (optional – see Note 8). In this case, 3-prime ends of primers are all oriented towards the unknown 5-prime region of mRNA to be elucidated.

2.2.2. Terminal Transferase-Mediated 5′ RACE LaNe (TT RACE LaNe)

1. 200 ng total RNA (see Note 1).
2. 10 mM dNTPs.
3. SuperScript™ II Reverse Transcriptase and accompanying buffers (Invitrogen) (see Note 9).
4. 25 mM manganese chloride.
5. 5 mg/ml bovine serum albumin.
6. RT primer: Reverse transcription primer should be oriented such that the 3-prime end is directed toward the unknown 5-prime mRNA region to be elucidated and reverse transcribed product includes features "1" to "6" of Fig. 1.
7. RNaseH (Invitrogen).
8. QIAEXII DNA purification system (Qiagen).

9. Terminal transferase (New England Biolabs).
10. 10 mM dATP.
11. *PfuTurbo*® DNA Polymerase and accompanying buffers and solutions (Stratagene) (see Note 3).
12. GSPT12G3 primer: 5′-gene-specific sequence followed by TTTTTTTTTTTTGGG-3′. The 5-prime gene-specific region of this primer should be in the same orientation as GSP1 and GSP4 (see Subheading 2.2.2, item 13) in alignment with position "3" as depicted in Fig. 1.
13. GSP1 and GSP4: Gene-specific primers 1 and 4 (see Subheading 2.1, item 7 and Fig. 1). In this case, 3-prime ends of primers are all oriented towards the unknown 5-prime region of mRNA to be elucidated.
14. GSP2 and GSP5: Gene-specific primers 2 and 5 (see Subheading 2.1, item 8 and Fig. 1). In this case, 3-prime ends of primers are all oriented towards the unknown 5-prime region of mRNA to be elucidated.
15. GSP6: Gene-specific primer 6 (see Subheading 2.1, item 10 and Fig. 1) (optional – see Note 8). In this case, 3-prime ends of primers are all oriented towards the unknown 5-prime region of mRNA to be elucidated.

2.3. LaNe RAGE

1. 300 ng genomic DNA (see Note 1).
2. 10 mM dNTPs.
3. *PfuTurbo*® DNA Polymerase and accompanying buffers and solutions (Stratagene) (see Note 3).
4. NNN3: First-strand synthesis primer 5′-gene-specific sequence followed by NNNNNCTCAC-3′ (see Note 10). The 5-prime gene-specific region of this primer should be in the same orientation as GSP1 and GSP4 (see Subheading 2.3, item 5) in alignment with position "3" as depicted in Fig. 1.
5. GSP1 and GSP4: Gene-specific primers 1 and 4 (see Subheading 2.1, item 7 and Fig. 1). In this case, 3-prime ends of primers are all oriented towards the unknown region of genomic DNA to be elucidated.
6. GSP2 and GSP5: gene-specific primers 2 and 5 (see Subheading 2.1, item 8 and Fig. 1). In this case, 3-prime ends of primers are all oriented towards the unknown region of genomic DNA to be elucidated.
7. QIAEXII DNA purification system (Qiagen) (optional – see Note 8).
8. GSP6: Gene-specific primer 6 (see Subheading 2.1, item 10 and Fig. 1) (optional – see Note 8). In this case, 3-prime ends of primers are all oriented towards the unknown region of genomic DNA to be elucidated.

3. Methods

LaNe-based methods represent a suite of tools that are applicable to the elucidation of 5-prime and 3-prime mRNA flanking sequence (5′RACE LaNe and 3′RACE LaNe, respectively) and genomic DNA flanking sequence in either direction (LaNe RAGE). Theoretically, the latter could be applied to the determination of insertion elements such as transposons as part of mutagenesis screening or DNA fingerprinting strategies, although these applications remain to be tested. The examples that follow represent demonstration methods that have been tested. It should be noted that elements of the respective approaches could be applicable to alternative approaches within the suite of tools (see Notes 11 and 12).

3.1. 3′RACE LaNe

The reaction scheme for 3′RACE LaNe is similar to that depicted in Fig. 1 for LaNe RAGE, except NNN3 is swapped for the reverse transcription primer TTT3 in the first step.

1. TTT3 primer (1 μM final concentration) is used to prime ThermoScript™-catalysed reverse transcription of 20 ng total RNA at 55°C for 1 h in a total reaction volume of 20 μl.
2. Reverse transcription products are treated with 2 U RNaseH at 37°C for 20 min prior to heat inactivation at 85°C for 10 min.
3. Five microliters of this product are used to template a 50 μl *PfuTurbo*® hot start PCR in (see Note 13) 1× reaction buffer including 300 μM dNTPs, 3.75 U *PfuTurbo*® DNA Polymerase, and 250 nM GSP1 and GSP4. Thermocycling parameters used are: 40 cycles of 94°C for 1 min, 50°C for 1 min, 72°C for 5 min (see Note 7).
4. Product can be analyzed by agarose gel electrophoresis at this stage (see Note 6). Single band or multiple banding profiles might be expected depending on experimental conditions (see Note 14). Optionally, a nested PCR can be performed, in which case, proceed to the next step.
5. Use 0.5 μl of product from the preceding step to template a nested 50 μl *PfuTurbo*® hot start PCR in (see Note 13) 1× reaction buffer including 300 μM dNTPs, 3.75 U *PfuTurbo*® DNA Polymerase, and 250 nM GSP2 and GSP5. Thermocycling parameters used are: 40 cycles of 94°C for 1 min, 58°C for 1 min, 72°C for 5 min (see Note 7).
6. Product can be analyzed by agarose gel electrophoresis at this stage. Single band or multiple banding profiles might be expected depending on experimental conditions (see Note 14).
7. Product can be subjected to direct Sanger sequencing using GSP5 or GSP6 as the sequencing primer (see Note 8).

8. Selected post-agarose gel electrophoresis resolved products can be gel-extracted using the QIAEXII gel extraction system, followed by direct Sanger sequencing using GSP5 or GSP6 as the sequencing primer (see Note 8).

3.2. 5′ RACE LaNe

Two example protocols are presented: FI RACE LaNe and TT RACE LaNe. The former uses a relatively streamlined workflow but a less "defined" mechanism for lariat structure formation (see Note 2). As such, FI RACE LaNe may be useful in some contexts but not others. The first step of FI RACE LaNe employs a non-hybrid reverse transcription primer instead of NNN3 depicted in Fig. 1 and relies on fortuitous lariat structure formation and extension by the first DNA strand. Ensure that features 1–6 with reference to Fig. 1 are oriented such that feature 6 is closest to the unknown 5-prime sequence. TT RACE LaNe is designed to control lariat structure formation to facilitate a more robust approach but uses a more cumbersome protocol to achieve this. Similarly, ensure that features 1 to 6 are oriented such that feature 6 is closest to the unknown 5-prime sequence. Reverse transcription and terminal transferase processing precede priming by GSPT12G3, which is analogous to NNN3 priming depicted in Fig. 1.

3.2.1. FI RACE LaNe

1. RT primer (1 μM final concentration) is used to prime ThermoScript™-catalyzed reverse transcription of 200 ng total RNA at 50°C for 1 h in a total reaction volume of 20 μl.
2. Reverse transcription products are treated with 2 U RNaseH at 37°C for 20 min prior to heat inactivation at 85°C for 10 min.
3. Two microliters of this product are used to template a 50 μl *PfuTurbo*® hot start PCR in (see Note 13) 1× reaction buffer including 300 μM dNTPs, 3.75 U *PfuTurbo*® DNA Polymerase, and 250 nM GSP1 and GSP4. Thermocycling parameters used are: 40 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 2 min (see Note 7).
4. Product can be analyzed by agarose gel electrophoresis at this stage (see Note 6). Single band or multiple banding profiles might be expected depending on experimental conditions (see Note 14). Optionally, a nested PCR can be performed, in which case, proceed to the next step.
5. Use 0.5 μl of product from the preceding step to template a nested 50 μl *PfuTurbo*® hot start PCR in (see Note 13) 1× reaction buffer including 300 μM dNTPs, 3.75 U *PfuTurbo*® DNA Polymerase, and 250 nM GSP2 and GSP5. Thermocycling parameters used are: 40 cycles of 94°C for 1 min, 60°C for 1 min, 72°C for 2 min (see Note 7).
6. Product can analyzed by agarose gel electrophoresis at this stage. Single band or multiple banding profiles might be expected depending on experimental conditions (see Note 14).

7. Product can be subjected to direct Sanger sequencing using GSP5 or GSP6 as the sequencing primer (see Note 8).
8. Selected post-agarose gel electrophoresis resolved products can be gel-extracted using the QIAEXII gel extraction system, followed by direct Sanger sequencing using GSP5 or GSP6 as the sequencing primer (see Note 8).

3.2.2. TT RACE LaNe

1. RT primer (1 μM final concentration) is used to prime SuperScript™ II-catalyzed reverse transcription of 200 ng total RNA in 1× buffer in the presence of 2 mM manganese chloride and 0.1 mg/ml bovine serum albumin at 42°C for 1 h in a total reaction volume of 20 μl (see Note 9).
2. Reverse transcription products are treated with 2 U RNaseH at 37°C for 20 min prior to heat inactivation at 85°C for 10 min.
3. Products are cleaned using the QIAEXII DNA purification system and eluted using 20 μl of elution buffer. 15 μl of eluate is collected, with care taken to avoid transfer of matrix slurry to the collection tube.
4. Five microliters of eluate is included in a 20 μl terminal transferase reaction in 1× buffer with the inclusion of 200 μM dATP and 20 U terminal transferase at 37°C for 20 min.
5. Two microliters of this product are used to template a 50 μl *PfuTurbo*® hot start PCR in (see Note 13) 1× reaction buffer including 300 μM dNTPs, 3.75 U *PfuTurbo*® DNA Polymerase, 100 nM GSPT12G3, and 250 nM GSP1 and GSP4. Thermocycling parameters used are: 40 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 2 min (see Note 7).
6. Product can be analyzed by agarose gel electrophoresis at this stage (see Note 6). Single band or multiple banding profiles might be expected depending on experimental conditions (see Note 14). Optionally, a nested PCR can be performed, in which case, proceed to the next step.
7. Use 0.5 μl of product from the preceding step to template a nested 50 μl *PfuTurbo*® hot start PCR in (see Note 13) 1× reaction buffer including 300 μM dNTPs, 3.75 U *PfuTurbo*® DNA Polymerase, and 250 nM GSP2 and GSP5. Thermocycling parameters used are: 40 cycles of 94°C for 1 min, 60°C for 1 min, 72°C for 2 min (see Note 7).
8. Product can analyzed by agarose gel electrophoresis at this stage. Single band or multiple banding profiles might be expected depending on experimental conditions (see Note 14).
9. Product can be subjected to direct Sanger sequencing using GSP5 or GSP6 as the sequencing primer (see Note 8).
10. Selected post-agarose gel electrophoresis resolved products can be gel-extracted using the QIAEXII gel extraction

system, followed by direct Sanger sequencing using GSP5 or GSP6 as the sequencing primer (see Note 8).

3.3. LaNe RAGE

Figure 1 schematically depicts the steps in LaNe RAGE. It must be ensured that the orientation of features 1–6 are such that feature 6 is closest to the unknown region to be characterized.

1. 50 µl 1× *PfuTurbo*® buffer containing 300 ng genomic DNA and 250 nM NNN3 is heated to 94°C for 3 min and then placed on ice for 5 min prior to the inclusion of 3.75 U *PfuTurbo*® and dNTPs to achieve a concentration of 300 µM.
2. Heat the reaction tubes from 18°C to 72°C at a rate of 0.1°C per 5 s and hold at 72°C for 10 min prior to placing on ice (see Note 10).
3. Include GSP1 and GSP4 to achieve 250 nM concentrations of each and proceed immediately to thermocycling. Thermocycling parameters used are: 40 cycles of 94°C for 1 min, 58°C for 1 min, 72°C for 2 min (see Note 7).
4. Product can be analysed by agarose gel electrophoresis at this stage (see Note 6). Single band or multiple banding profiles might be expected depending on experimental conditions (see Note 7). Optionally, a nested PCR can be performed, in which case, proceed to the next step.
5. Use 0.5 µl of product from the preceding step to template a nested 50 µl *PfuTurbo*® hot start PCR in (see Note 13) 1× reaction buffer including 300 µM dNTPs, 3.75 U *PfuTurbo*® DNA Polymerase, and 250 nM GSP2 and GSP5. Thermocycling parameters used are: 40 cycles of 94°C for 1 min, 60°C for 1 min, 72°C for 2 min (see Note 7).
6. Product can analysed by agarose gel electrophoresis at this stage. Single band or multiple banding profiles might be expected depending on experimental conditions (see Note 14).
7. Product can be subjected to direct Sanger sequencing using GSP5 or GSP6 as the sequencing primer (see Note 8).
8. Selected post-agarose gel electrophoresis resolved products can be gel-extracted using the QIAEXII gel extraction system, followed by direct Sanger sequencing using GSP5 or GSP6 as the sequencing primer (see Note 8).

4. Notes

1. The integrity of the input nucleic acid specimen will have a considerable bearing on the potential for success using LaNe-based approaches, as with most molecular biology technologies. If the input material is significantly degraded or fragmented, only

relatively short regions of flanking sequence may be resolvable in any given "LaNe step." In this context, it is perhaps unwise to design primers and conditions in an attempt to yield long products.

2. ThermoScript™ Reverse Transcriptase is relatively thermostable, permitting relatively stringent priming of reverse transcription, if required. Alternative enzymes with similar properties should also be suitable.
3. Terminal transferase activity is undesirable in a thermostable DNA-dependent DNA polymerase employed in LaNe-based methods in which the lariat structure formation is facilitated by a gene-specific hybrid primer sequence incorporation event, e.g., via TTT3, GSPT12G3, or NNN3. Such activity would result in a majority of would-be lariat-primed events exhibiting mismatches at the 3-prime end, thus inhibiting lariat-primed extension dramatically. This does not apply to examples such as FI RACE LaNe or in other cases in which the principle of "fortuitous" lariat structure priming is applied (e.g., in principle, an iteration of LaNe RAGE could remove the requirement for NNN3-primed strand synthesis to simply employ nested PCR using GSP1 and GSP4, followed by GSP2 and GSP5 – see Notes 6 and 12). *PfuTurbo*® DNA Polymerase has been used effectively, but alternative enzymes with similar properties should also suit LaNe-based methods.
4. In the LaNe-based methods, for all primers used, gene-specific primer regions are always oriented with the 3-prime end towards the unknown sequence to be elucidated. In cases where hybrid primer is used, e.g., TTT3, GSPT12G3, and NNN3, the gene-specific region is always located at the 5-prime end of the primer, but the orientation of the gene-specific primer sequence is designed with its 3-prime end oriented towards the unknown sequence to be elucidated as with all other primers. This works because end-filling and subsequent lariat-based priming (as part of the LaNe mechanism) yield reverse complement forms of some sites at one end of the template, allowing primers to engage in two-sided PCR.
5. Primer design software such as OligoAnalyzer 3.1, Primer 3, and Primer-BLAST can be employed to design primers according to the desired melting temperature (aim for "typical" PCR primer properties) and to check for and avoid primers or primer combinations liable to form strong secondary structures (e.g., hairpins) or engage in primer dimer formation. PCR primers can be filtered against repetitive elements of reference genomes and checked for predicted off-target products. The scope for primer design is somewhat dependent, however, on the amount of known sequence available and the desire to design the primers to anneal close to the unknown region to be resolved if possible.

6. In some applications, just one round of PCR may be sufficient. This requires fewer primers and a shorter region of known sequence to design LaNe primers. There will be other cases where a single round of PCR is insufficient, and the added sensitivity and specificity afforded by a second nested PCR will resolve desired products.
7. The RT and PCR conditions detailed in this chapter represent proof-of-concept methodologies. These are most likely open to considerable optimization to suit requirements. For example, some applications may require greater or lower stringency necessitating higher or lower annealing temperatures, respectively. Some applications may require longer extension times during cycling to facilitate yield of longer products. Applications may require more or less enzyme or initial nucleic acid template. It is likely that fewer PCR cycles would be necessary than are detailed in this chapter. However, it should be noted that by their nature, LaNe-based protocols employ templates that form suppressive "panhandle" structures. As such, relatively high concentrations of primers GSP1 and GSP4 or GSP2 and GSP5 may be required to compete effectively with such structure formation for effective priming during thermocycling. Ideally, the primer regions should be relatively close together to minimize the relative stability and rate of formation of such structure formation (primer sites could even overlap to a limited extent). Conceivably, GSP1 and GSP4 or GSP2 and GSP5 could employ "heel clamps" to increase Tm and improve competitivity.
8. As a fast and convenient approach, products can be sequenced directly using GSP5 or GSP6 as the sequencing primer. In some cases, it may be desirable or necessary to isolate product bands from an agarose gel. Following gel extraction, products can be separately sequenced using GSP5 or GSP6 as the sequencing primer. Ideally, sequencing primers should be set back sufficiently from the unknown flanking sequence under study to allow verification of gene-specific amplicon. If the former is not possible due to design constraints or if preferred, product can be cloned into a plasmid and sequencing performed using plasmid-specific sequencing primers.
9. TT RACE LaNe makes use of the inherent terminal transferase activity of M-MLV-based reverse transcriptase, e.g., SuperScript™ II or Expand-RT, in the presence of manganese chloride as an additional divalent cation to magnesium chloride and in the presence of BSA as a stabilizing agent. Under these conditions, several cytosine residues are typically added in the presence of 5′ capped mRNA templates. In the presence of 5′-OH or 5′-phosphate termini, the activity is less pronounced and not cytosine-specific (3).

10. As a proof-of-concept utility, NNN3 was designed with a "NNNNNCTCAC" 3-prime terminus. However, any number of variations could be used to bias towards more frequent or less frequent hybridization sites to influence the tendency of product sizes. For instance, a longer "defined" region at the 3-prime end might tend towards the yield of longer products, since hybrid primer binding sites would be less frequent. The nature of the defined sequence at the 3-prime end could be varied to suit the experimental context. There is considerable scope to influence the performance of the system by adjusting hybrid primer sequence. Integrally related to this is the way in which annealing of NNN3 to denatured genomic DNA is conducted prior to extension. More stringent conditions should result in fewer hybrid primer binding events and fewer products, to enable the yield of longer products. For example, immediately following denaturation (in the presence of template, NNN3, *PfuTurbo*® DNA Polymerase, and dNTPs in reaction buffer) one might perform annealing at an elevated temperature, e.g., 45–55°C, for several minutes (without having cooled below this temperature) prior to slowly raising the temperature to 72°C (e.g., 1°C/20 s to balance proficiency of extension at suboptimal temperatures with maintenance of annealing), with all subsequent primers added at this temperature. Further, a reduction in concentration of NNN3 may contribute to less frequent hybrid primer binding events to permit "cleaner" product profiles and selection of longer products.
11. It is likely that LaNe RAGE would be useful in applications such as transposon insertional mutagenesis screening programs to identify insertion sites of mutants and in DNA fingerprinting, as has been demonstrated using Universal Fast Walking. However, these experiments have not been conducted.
12. One approach that may be valuable but has not been tested to date would combine aspects of CapTrapper (16) and 5′RACE LaNe, in which 5-prime capped mRNAs would be physically separated from noncapped mRNA species prior to utilization of a LaNe-based approach. Fortuitous lariat structure formation might be employed in LaNe RAGE without first-strand synthesis primed by NNN3, analogous to that used in TT RACE LaNe. This would result in a less-controlled but more streamlined protocol in which the user could proceed directly to a hot-start PCR templated with genomic DNA using primers GSP1 and GSP4, optionally followed by a nested PCR using GSP2 and GSP5.
13. In the proof-of-concept methods detailed, hot-start PCR was performed by separate addition of enzyme following thermal

denaturation. Alternatively (and preferably), chemical or antibody inhibition-based hot-start enzyme may be employed.

14. Product profiles vary considerably and are strongly influenced by hybridization conditions, primer design, and PCR parameters as discussed in this section.

Acknowledgments

This work was supported by the National Health and Medical Research Council of Australia grant no. 148630 to Professors J.A.M. Graves and M.B. Renfree, Associate Professor V.R. Harley and Dr. A.J. Pask.

References

1. Frohman, M.A., Dush, M.K., and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**, 8998–9002.
2. Frohman, M.A. and Martin, G.R. (1989) Rapid amplification of cDNA ends using nested primers. *Technique* **1**, 165–70.
3. Schmidt, W.M. and Mueller, M.W. (1999) CapSelect: a highly sensitive method for 5′CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs. *Nucleic Acids Res.* **27**, e31.
4. Maruyama, K. and Sugano, S. (1994) Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* **138**, 171–4.
5. Volloch, V., Schweitzer, B., and Rits, S. (1994) Ligation-mediated amplification of RNA from murine erythroid cells reveals a novel class of beta globin mRNA with an extended 5′-untranslated region. *Nucleic Acids Res.* **2**, 2507–11.
6. Cormack, R.S. and Somssich, I.E. (1997) Rapid amplification of genomic ends (RAGE) as a simple method to clone flanking genomic DNA. *Gene* **194**, 273–6.
7. Ochman, H., Gerber, A.S., and Hartl, D.L. (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621–3.
8. Mueller, P.R. and Wold, B. (1989) In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246**, 780–6.
9. Devon, R.S., Porteous, D.J., and Brookes, A.J. (1995) Splinkerettes – improved vectorettes for greater efficiency in PCR walking. *Nucleic Acids Res.* **23**, 1644–5.
10. Liu, Y.-G. and Whittier, R.F. (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* **25**, 674–81.
11. Myrick, K.V. and Gelbart, W.M. (2002) Universal fast walking for direct and versatile determination of flanking sequence. *Gene* **284**, 125–31.
12. Tonooka, Y. and Fujishima, M. (2009) Comparison and critical evaluation of PCR-mediated methods to walk along the sequence of genomic DNA. *Appl. Microbiol. Biotechnol.* **85**, 37–43.
13. Park, D.J. (2004) 3′RACE LaNe: a simple and rapid fully nested PCR method to determine 3′-terminal cDNA sequence. *Biotechniques* **36**, 586–90.
14. Park, D.J. (2005) A new 5′ terminal murine GAPDH exon identified using 5′RACE LaNe. *Mol. Biotechnol.* **29**, 39–45.
15. Park, D.J. (2005) LaNe RAGE: a new tool for genomic DNA flanking sequence determination. *Electronic J. Biotechnol.* **8**, 218–25.
16. Carninci, P., Westover, A., Nishiyama, Y., Ohsumi, T., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Murumatsu, M., Schneider, C., and Hayashizaki, Y. (1997) High efficiency selection of full-length cDNA by improved biotinylated cap trapper. *DNA Res.* **4**, 61–6.

Chapter 5

CODEHOP PCR and CODEHOP PCR Primer Design

Jeannette P. Staheli, Richard Boyce, Dina Kovarik, and Timothy M. Rose

Abstract

While PCR primer design for the amplification of known sequences is usually quite straightforward, the design, and successful application of primers aimed at the detection of as yet unknown genes is often not. The search for genes that are presumed to be distantly related to a known gene sequence, such as homologous genes in different species, paralogs in the same genome, or novel pathogens in diverse hosts, often turns into the proverbial search for the needle in the haystack. PCR-based methods commonly used to address this issue involve the use of either consensus primers or degenerate primers, both of which have significant shortcomings regarding sensitivity and specificity. We have developed a novel primer design approach that diminishes these shortcomings and instead takes advantage of the strengths of both consensus and degenerate primer designs, by combining the two concepts into a Consensus–Degenerate Hybrid Oligonucleotide Primer (CODEHOP) approach. CODEHOP PCR primers contain a relatively short degenerate 3′ core and a 5′ nondegenerate clamp. The 3′ degenerate core consists of a pool of primers containing all possible codons for a 3–4 aminoacid motif that is highly conserved in multiply aligned sequences from known members of a protein family. Each primer in the pool also contains a single 5′ nondegenerate nucleotide sequence derived from a codon consensus across the aligned aminoacid sequences flanking the conserved motif. During the initial PCR amplification cycles, the degenerate core is responsible for specific binding to sequences encoding the conserved aminoacid motif. The longer consensus clamp region serves to stabilize the primer and allows the participation of all primers in the pool in the efficient amplification of products during later PCR cycles. We have developed an interactive web site and algorithm (iCODEHOP) for designing CODEHOP PCR primers from multiply aligned protein sequences, which is freely available online. Here, we describe the workflow of a typical CODEHOP PCR assay design and optimization and give a specific implementation example along with "best-practice" advice.

Key words: PCR, CODEHOP, Consensus, Degenerate, Acyl-coA binding protein

1. Introduction

The development of PCR-based assays to identify unknown distantly related genes or pathogens is problematic and relies upon mixtures of nucleic acid primers and the ability of primers

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_5, © Springer Science+Business Media, LLC 2011

to hybridize to noncomplementary sequences with a required degree of specificity. Pools of related primers carrying known or predicted nucleotide sequence differences throughout the length of the primer have been used with moderate success to amplify unknown or distantly related genes. These are referred to as degenerate primers and can contain hundreds or thousands of individual primers in the pool to cover all possible nucleotide variations in a particular sequence. Alternatively, consensus PCR primers have also been utilized to amplify unknown or related sequence variants. A consensus primer carries the most common actual or predicted nucleotide variant in each position of a primer sequence and relies on its ability to specifically hybridize to a target sequence with mismatched or unpaired bases. When basing primer design on protein coding sequences, standard degenerate primers will contain most or all of the possible nucleotide sequences encoding a large conserved aminoacid motif, while consensus primers will contain the most common nucleotide at each codon position in the targeted motif. While useful with adequate concentrations of closely related template targets in noncomplex mixtures, both standard degenerate- and consensus-primer approaches suffer from a lack of specificity and sensitivity when these conditions are not met.

We have developed a PCR approach for detecting and identifying unknown and distantly related gene sequences using consensus–degenerate hybrid oligonucleotide primers (CODEHOPs) (1–3) (Fig. 1). CODEHOPs are designed from short highly conserved motifs identified in multiply aligned protein sequences from members of a gene family and are used in PCR amplification to identify unknown members of the family. Each CODEHOP consists of a short 3′ degenerate core region corresponding to all possible codons specifying 3–4 highly conserved amino acids and a longer 5′ consensus clamp region containing a single "best guess" nucleotide sequence derived from the consensus sequences flanking the target motif. Thus, a CODEHOP PCR primer consists of a pool of primers that are heterogeneous at the 3′ end and homogeneous at the 5′ end.

The CODEHOP primer design strategy overcomes problems of both degenerate and consensus PCR primer methods. The limited degeneracy in the short 3′ core region minimizes the total number of individual primers in the degenerate pool yet provides a broad specificity during the initial PCR amplification cycles. Hybridization of the 3′ degenerate core is stabilized by the 5′ consensus clamp, which allows higher annealing temperatures without increasing the degeneracy of the primer pool. Although mismatches between the 5′ consensus clamp and the target sequence may occur during the initial PCR cycles, they are situated away from the 3′ hydroxyl extension site of the polymerase, thus minimizing their disruptive effects on polymerase priming

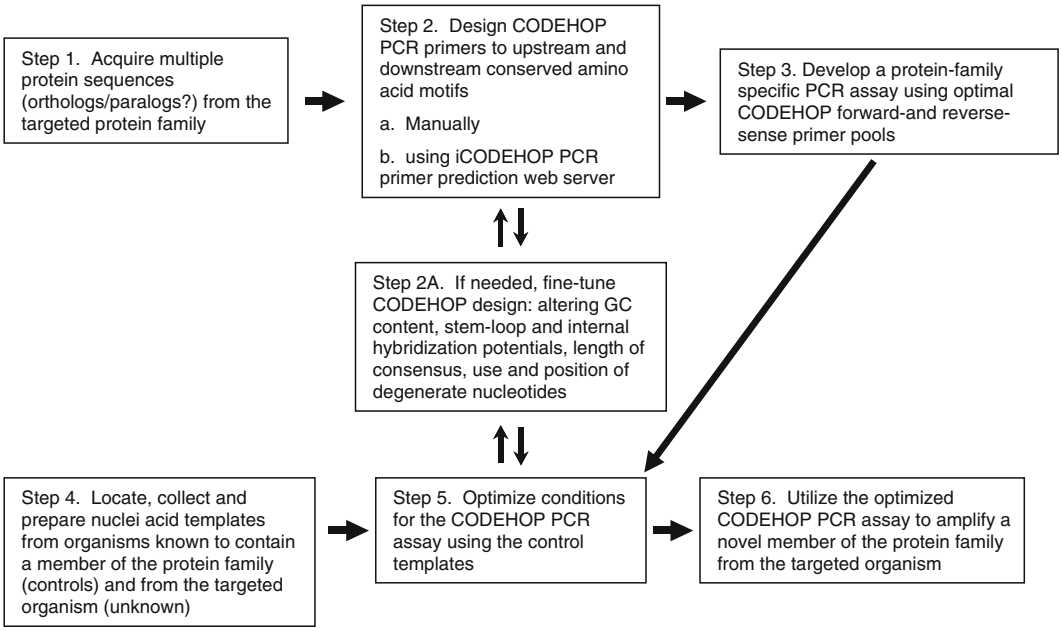


Fig. 1. CODEHOP PCR assay development. The general flowchart for development of CODEHOP PCR assays is shown. The different steps are described in the text.

and extension. Further amplification of primed PCR products during subsequent rounds of primer hybridization and extension is enhanced by the sequence similarity of all primers in the pool. This allows utilization of all primers in the PCR reaction cycles, and a shortage of the one or a few specific primers does not become a limiting factor. The CODEHOP PCR approach provides the necessary specificity and sensitivity to allow for the amplification of distantly related genetic homologs or paralogs in diverse species or disparate pathogen species, at low titer, in complex mixtures of genetic material (4–15).

The following protocol describes the general method for the design and use of CODEHOP PCR primers for the amplification of novel genes, with a specific example targeting the acyl-CoA-binding protein (ACBP) family. The ACBP family is involved in multiple essential cellular tasks including modulation of fatty acid biosynthesis, enzyme regulation, regulation of the intracellular acyl-CoA pool size, donation of acyl-CoA esters for β-oxidation, vesicular trafficking, complex lipid synthesis, and gene regulation (16). ACBP homologs have been identified in all four eukaryotic kingdoms, Animalia, Plantae, Fungi, and Protista, and in 11 eubacterial species. Using ACBP-specific CODEHOP primers similar to the ones used in the example below, we discovered a novel ACBP gene in yeast (4) and identified an ACBP pseudogene in humans (5). In 2005, we published an extensive analysis of ACBP sequences from over 50 different species (16).

The ACBP protein is highly conserved across phylums and a number of species, ranging from protozoa to vertebrates, and have evolved two to six lineage-specific paralogs through gene duplication and/or retrotransposition events. In a recent collaboration with the Northwest Association for Biomedical Research (NWABR), we have designed CODEHOP PCR primers to detect novel ACBPs in diverse species of plants. In this project, the NWABR partnered with high-school teachers to develop the BIO-ITEST curriculum, which was designed to help secondary school teachers and their students learn how information technology is used in biological research (17). Below, the design and implementation of the ACBP-specific CODEHOP primers targeted at plants is used as an example. However, the same concept can be applied to the discovery of distantly related unknown members of any other gene family of interest.

2. Materials

2.1. Software and Databases Used for Primer Design

1. NCBI BLAST suite (http://blast.ncbi.nlm.nih.gov/Blast.cgi).
2. NCBI Protein Database (http://www.ncbi.nlm.nih.gov/protein).
3. NCBI Nucleotide Database (http://www.ncbi.nlm.nih.gov/nucleotide).
4. ClustalW multiple sequence alignment software (e.g., http://www.ch.embnet.org/software/ClustalW.html).
5. iCODEHOP primer design software (https://icodehop.cphi.washington.edu/) (3).

2.2. Template DNA Preparation

1. DNeasy Plant Mini Kit (Qiagen) (see Note 1).
2. Cells or tissue from which DNA will be extracted.
3. Proteinase K digestion: 50 µg/ml Proteinase K in 100 µg/ml in 0.01 M Tris–HCl, pH 7.8, 5 mM EDTA, 0.5% SDS.
4. Saturated phenol–chloroform–isoamyl ethanol, 25:24:1, stored at 4°C. Phenol and its vapors are highly corrosive; wear gloves and work in a fume hood when handling. Phenol dissolves polystyrene plastics, so glass or polypropylene pipettes and tubes should be used.
5. Chloroform ($CHCl_3$) is a CNS depressant and a suspected carcinogen, so exposure to vapors should be avoided.
6. 100% Ethanol.
7. TE buffer: 10 mM Tris–HCl, pH 8.0, 1 mM EDTA.

2.3. PCR Analysis

2.3.1. PCR Master Mix (see Note 2)

1. 10× PCR buffer: 200 mM Tris–HCl (pH 8.4), 500 mM KCl (Invitrogen).
2. $MgCl_2$ stock solution, 25 nM (Invitrogen).
3. PCR-grade nucleotide mix: 10 mM dNTP mix (Fermentas).
4. Platinum Taq DNA polymerase, 5 U/μl (Invitrogen).
5. PCR-grade water.

2.3.2. Oligonucleotides/DNA

1. Forward and reverse CODEHOP primers are dissolved at 50 μM in nuclease-free water. Oligonucleotides are stable for several months at –20°C. For long-term storage of stocks, samples can be dried down and kept at –20°C.
2. Template DNA, stored at 4°C in TE solution or at –20°C in nuclease-free water.

2.3.3. Gel Electrophoresis

1. DNA-grade agarose: GenePure LE (ISC BioExpress); 1.5 g of agarose is dissolved in 75 ml TAE running buffer, heated in a microwave to dissolve the agarose, and poured into a gel box to create a 2% gel with space to run 32 samples plus DNA ladders to size the amplicons.
2. Loading buffer: 6× loading buffer (Fermentas).
3. Running buffer: 40 mM Tris–acetate, 1 mM EDTA (TAE buffer).
4. DNA ladders: 1 kb GeneRuler (Fermentas).
5. Ethidium bromide: 0.5 μg/ml ethidium bromide in water.
6. UV imaging: BioRad UV box to image the bands on gels.

3. Methods

The search for a new gene using the CODEHOP approach consists of four steps. First, CODEHOP PCR primers are designed manually or using our recently improved interactive iCODEHOP design web site and algorithm (3). In both cases, online databases and software tools are used to identify protein family members, to download protein and nucleotide sequences, to perform sequence alignments, and to identify conserved sequence blocks from which CODEHOP PCR primers are derived. The second task involves the identification and preparation of suitable sources of control and targeted nucleic acid templates. The third step includes the actual CODEHOP PCR assay and optimization. If necessary, the primer design can be adjusted. The fourth step includes sequencing of amplicons followed by sequence and phylogenetic analysis.

3.1. Primer Design: Manual Method

1. Collect the protein sequences from known members of the targeted gene family. Perform a BLAST search in the NCBI protein database using the protein sequence of a member of the targeted gene family. In our example, the ACBP sequence from *Arabidopsis thaliana* (NP_174462) was used as query in a BLAST search that yielded ten additional complete plant ACBP sequences (see Fig. 2).
2. Align the protein sequences of interest using the ClustalW multiple alignment program (18). The example in Fig. 2 shows an alignment of the plant ACBP sequences (see Note 3).
3. Identify a conserved N-terminal motif of ~4 amino acids that has limited codon degeneracy and is flanked by additional conserved residues upstream of that motif. In our example, a highly conserved N-terminal motif "Y/FKQA" was identified with restricted degeneracy that also was flanked upstream by additional conserved residues (highlighted in Fig. 2). Identify a second C-terminal conserved motif within 40–500 amino acids downstream of the first motif (see Note 4).

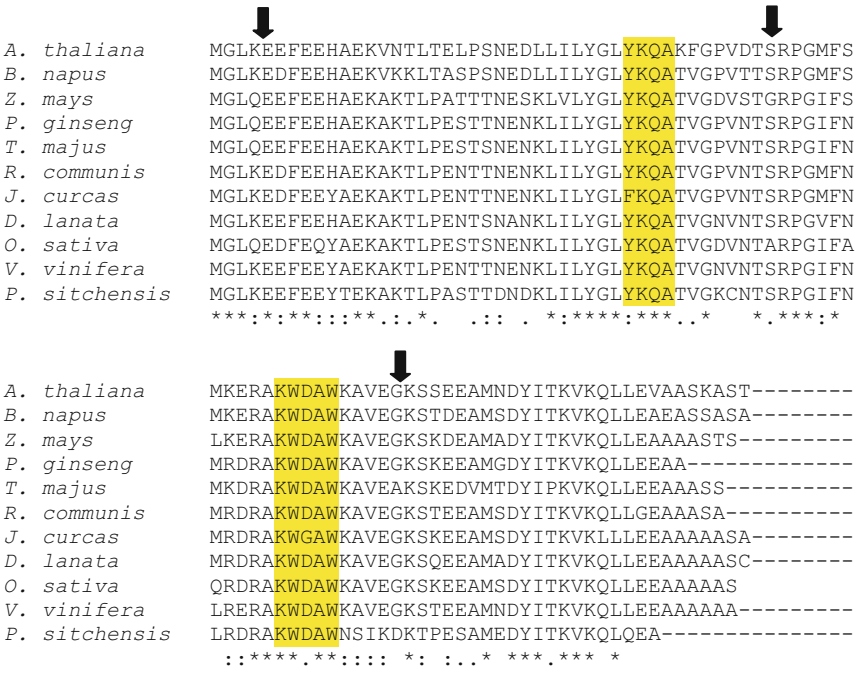


Fig. 2. Alignment of multiple protein sequences from the plant ACBP protein family targeted for CODEHOP PCR primer design. ACBP orthologs in plants were identified using a BLAST search of the NCBI protein database using the 92 aa *Arabidopsis thaliana* ACBP sequence as probe and aligned using ClustalW. Conserved aminoacid motifs "YKQA" and "KWDA" selected for CODEHOP PCR primer design are *highlighted*. Intron positions within the *A. thaliana* ACBP gene are indicated with an *arrow*. The sequences and their corresponding GenBank accession numbers are: *A. thaliana* (NP_174462), *Brassica napus* (CAA54390), *Zea mays* (NP_001147418), *Panax ginseng* (BAB85987), *Tropaeolum majus* (AAP82942), *Ricinus communis* (CAA70200), *Jatropha curcas* (ABE72959); *Digitalis lanata* (CAB56693), *Oryza sativa* (japonica cultivar) (NP_001061062), *Vitis vinifera* (XP_002263421), *Picea sitchensis* (ABK25059).

In our example, a conserved C-terminal motif "KWDA" was identified 19 amino acids downstream of the "Y/FKQA" motif that also has restricted codon degeneracy due to the conserved tryptophan residue (highlighted in Fig. 2). Although the "Y/FKQA" and "KWDA" motifs are only separated by 19 amino acids, they are interrupted by an intron that would increase the size of the PCR product (see Fig. 2). Analysis of the intron size from the known *A. thaliana* and *O. sativa* ACBP genes reveals a range of 104–728 bp. Thus, amplification between primers derived from the "Y/FKQA" and "KWDA" motifs would yield a PCR product of 200–900 bp for plant genes with similar size introns (see Note 5).

4. To design the degenerate core of the "sense" strand CODEHOP PCR primer targeting the desired upstream motif, determine the possible nucleotide sequences encoding the conserved aminoacid motif. We recommend not exceeding a primer degeneracy of 32- or 64-fold unless there is strong flanking sequence conservation. In the current example, the degenerate nucleotide sequence 5′ TWY AAR CAR GC 3′ (using I.U.B designations for multiple nucleotides: W = A,T; Y = C,T; R = A,G) contains all possible codons for the upstream "Y/FKQA" motif (Fig. 3a). The wobble position of the alanine codons (GCN) was not used to reduce degeneracy. To design the consensus clamp of the "sense" strand CODEHOP PCR targeting the desired upstream motif, determine the most common aminoacid at each position of the 6–7 amino acids upstream of the targeted motif. The most frequently used codon for each of these amino acids for the targeted organism is determined. Alternatively, obtain the corresponding encoding nucleotide sequences from the NCBI database for each of the proteins aligned and choose the most common nucleotide at each position to obtain a complete primer of ~30–32 nucleotides. In our example, the nucleotide sequences encoding the "YKQA" motif and upstream flanking sequences were obtained from the NCBI database for each of the 11 plant ACBP proteins in the alignment (Fig. 3a). The most common nucleotide at each position of the six codons upstream of the "YKQA" motif were chosen to form the consensus clamp, yielding a clamp sequence of 5′ GCTTATTCTCTATGGACTCT 3′. The consensus clamp is positioned at the 5′ end of the primer and is shown in upper case. Thus, the proposed sense strand CODEHOP PCR forward primer, named YKQAa (the small case a denotes a sense strand primer), is a pool of 16 different primers, heterogeneous at the 3′ end and homogeneous at the 5′ end with the sequence 5′ GCTTATTCTCTATGGACTCTwyaarcargc 3′ (Fig. 3a).
5. To design the antisense-strand reverse CODEHOP primer, the same considerations are used. The degenerate core is

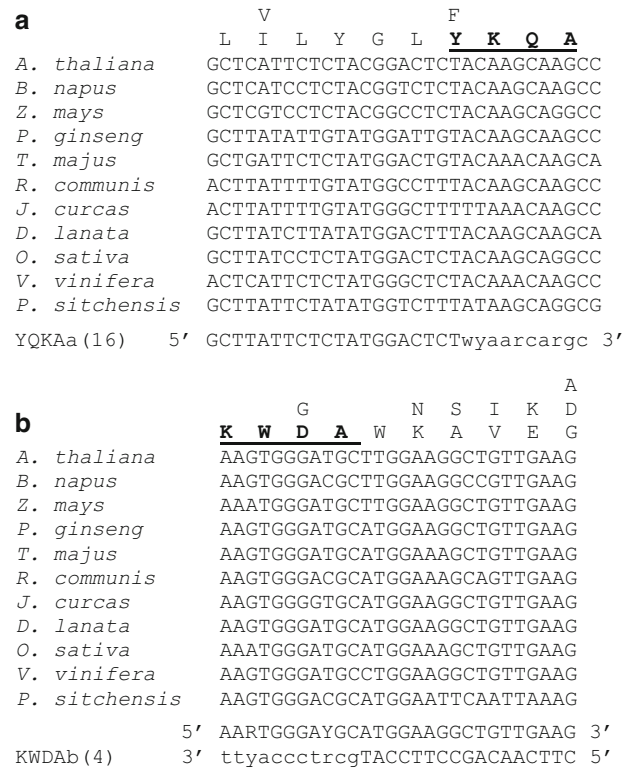


Fig. 3. Manual design of CODEHOP PCR primers targeting the conserved YKQA and KWDA conserved motifs of plant ACBP orthologs. The nucleotide sequences encoding the conserved (**a**) YKQA and (**b**) KWDA motifs and flanking regions from the plant ACBP genes identified by BLAST search were aligned, and the encoded aminoacid sequences are shown above. These aminoacid motifs were chosen for primer design due to the strong sequence conservation across a 4–5 aminoacid region, the presence of amino acids with restricted codons and regions of conserved flanking sequences. The degenerate cores of the CODEHOP PCR primers are shown in lower case using the I.U.B. code for degenerate nucleotides: Y = T,C; R = A,G; W = A,T. The nondegenerate consensus clamp regions are shown in upper case. (**a**) The CODEHOP PCR primer "YKQAa" designed from the YKQA motif is 16-fold degenerate and corresponds to the sense "coding" strand. (**b**) The CODEHOP PCR primer "KWDAb" designed from the KWDA motif is four-fold degenerate and corresponds to the antisense strand. The coding strand is shown for reference. The plant genes and corresponding GenBank accession numbers are *Arabidopsis thaliana* (NM_102916), Brassica napus (X77134), *Zea mays* (NM_001153946), *Panax ginseng* (AB071376), *Tropaeolum majus* (AY319307), *Ricinus communis* (Y08996), *Jatropha curcas* (DQ452088), *Digitalis lanata* (AJ249833), *Oryza sativa* (japonica cultivar) (NM_001057071), *Vitis vinifera* (XM_002263385), *Picea sitchensis* (EF085763).

determined from the possible nucleotide sequences encoding the four aminoacid downstream motif. In our example, the sequences encoding the "KWDA" motif were determined to be 5′ AAR TGG GAY GC 3′. The wobble position in the alanine codon GCN was not utilized to reduce degeneracy. The complementary sequence used for the reverse-strand

primer is 5′ GC RTC CCA YTT 3′ (Fig. 3b). The nonconserved glycine in the third position of the *Jatropha curcas* sequence is ignored in this case. The consensus clamp region of this primer can be determined from the most common codon for the sequences downstream of the motif. The actual sequences present in known members of the gene family can also be used, as described above. In our example, the nucleotide sequences encoding the "KWDA" motif and downstream flanking sequences were obtained for each of the 11 plant ACBP proteins and aligned (Fig. 3b). The consensus nucleotides at each position of the six codons downstream of the "KWDA" motif were chosen, yielding the sense-strand sequence of 5′ aartgggaygcATGGAAGGCTGTTGAAG″ 3′. The complement of this, i.e., 5′ CTTCAACAGCCTTCCATgcrtcccaytt 3′ gives the desired antisense-strand reverse CODEHOP primer called KWDAb (the small case b denotes an anti-sense strand reverse primer), as seen Fig. 3b. Thus, the KWDAWb reverse primer is a pool of four different primers, heterogeneous at the 3′ end and homogeneous at the 5′ end.

3.2. Primer Design: Using iCODEHOP Software

To aid in primer design, we have developed and recently improved a web-based site to predict CODEHOP PCR primers from blocks of conserved aminoacid sequences (3). The conserved sequence blocks are obtained from multiple related protein sequences from the targeted gene family. The sequence block output is linked directly to the iCODEHOP design software, which predicts and scores possible CODEHOP PCR primers from the different sequence blocks present in the protein of interest.

1. Initiate the iCODEHOP program at [https://icodehop.cphi.washington.edu/i-codehop-context/Welcome] and choose to run the program in a named session, saving data to the server, or in a nonnamed session. Select the "Design primers" option and enter or upload either protein sequences or protein sequence alignments at the prompt and proceed with analysis. In the first case, the program creates the alignment for the user using ClustalW. For our example, the 11 plant ACBP protein sequences were used as input, and a Clustal alignment was generated. Examine the alignment to confirm the expected sequence similarities. The program also provides a phylogenetic tree to determine the relatedness of the different input sequences. Provide a name for the clustal alignment and proceed. iCODEHOP will carve out conserved sequence blocks from the multiple alignment.
2. To design primers from the sequence blocks, accept the default parameters or provide new parameters to direct the output. Parameters include limitations on the degree of

degeneracy (default = 128), primer T_m (which determines primer length; default = 60), the use of different codon tables for design of the consensus clamp (default = *Homo sapiens*), and the number of primers being displayed. Other parameters are indicated, and their affect is described in the help files. In our example, the primer T_m default value of (60°C) was increased to 70°C to obtain a primer of approximately 30 nucleotides (see Note 6).

3. Select "Design Primers" and iCODEHOP and then use the BLOCKMAKER program to identify sequence blocks conserved within the input protein family members, and from these blocks, propose forward and reverse CODEHOP PCR primers for each of the sequence blocks. Choose primers that have low degeneracy, and they together with a proposed reverse primer will produce an amplicon of the desired length. In our example, a single large conserved sequence block (x7451vuxAA) of 85 amino acids was identified within the 87–92 aa input plant ACBP sequences (Fig. 4a). A consensus sequence was derived, with highly conserved residues labeled with an asterisk. Possible forward and reverse CODEHOP PCR primers are enumerated and displayed graphically, and are linked to the underlying multiple protein alignment (see Fig. 4b). Each primer is displayed as aligned to the protein block from which it has been derived, and values for degeneracy, ranges of predicted melting temperatures, and primer length are provided. The output includes a phylogenetic tree that allows the user to evaluate the phylogenetic relationship of the input protein sequences to aid in primer design (not shown) (see Note 7).
4. After choosing a forward primer, the program will then suggest a list of possible complementary primers, indicating the resulting amplicon length and the degree of predicted melting temperature overlap. Make sure that both primers have a reasonably low degeneracy (we prefer a degeneracy of 64 or lower) and overlapping T_m range which is given for each primer.
5. iCODEHOP designs the consensus clamp based on the codon for the aminoacid that is most common in each position of the alignment. The actual codon sequence used is the most preferred codon for that aminoacid for the targeted organism, taken from a codon usage table specified by the user. Alternatively, the user can determine the actual nucleotides used to encode the amino acids in the consensus region manually, as shown in the manual method above and utilize these sequences in the primer design. This step will be an integrated part of the next version of iCODEHOP.

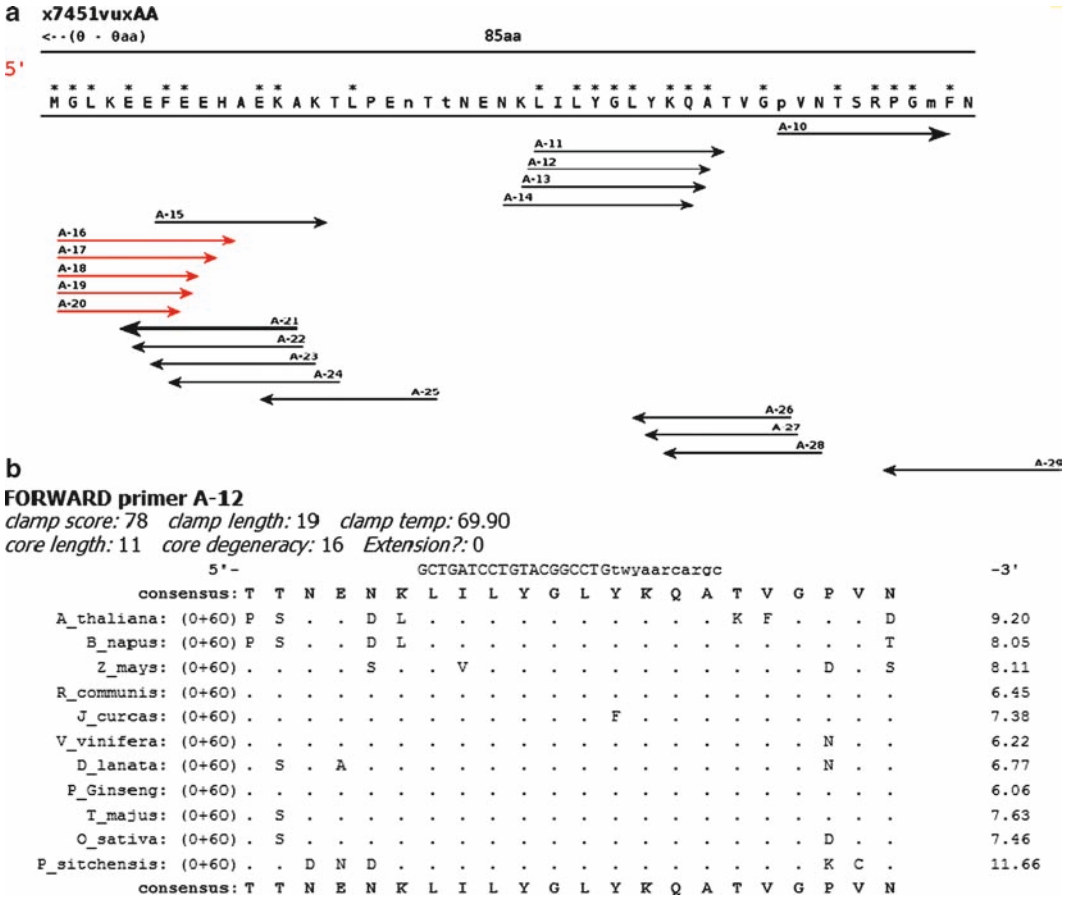


Fig. 4. Output of the iCODEHOP automated PCR primer design for the YKQA motif. The ACBP protein sequences from the diverse set of plants in Fig. 2 were used as input in the new interactive iCODEHOP web server to design CODEHOP PCR primers from the multiply aligned sequences. Primer design was performed using the default conditions, except for the use of a 68°C clamp melting temperature (discussed in the text). The portion of the graphical output of the analysis of a portion of the 85-aminoacid "(**a**)" block (x7451vuxAA) containing the YKQA motif is shown (*asterisks* indicate highly conserved residues). The *arrows* underneath the block indicate the length and position of possible sense- and antisense-strand CODEHOP primers, labeled with their block, and a number designation. The A-12 primer is derived from the complete YKQA motif. The predicted primer and the multiply aligned protein sequences that it is derived from [Shown in (**b**)] are linked to the A-12 primer *arrow* in (**a**). The information provided for the A-12 primer includes the length and melting temperature of the consensus clamp region, as well as the length and degeneracy of the core. *Dots* in the multiple alignment indicate identity with the consensus sequence. Further details regarding the output of the program are described on the iCODEHOP web site (19).

3.3. Template DNA Preparation

The source of template is dependent upon the targeted gene. In the present example, we were interested to identify novel ACBP genes in different plant species. For this purpose, DNA was isolated from leaves of various plants using a plant DNA isolation kit from Qiagen.

1. Isolate DNA from the tissue source of choice using standard procedures. Proteinase K extractions work well, although different isolation kits may be used.
2. Resuspend DNA in TE buffer at approximately 250 ng/µl.

3.4. CODEHOP PCR Amplification

CODEHOP PCR amplification can be performed using classical and touch-down approaches with a hot-start initiation (1). We recommend using a thermal gradient PCR amplification to empirically determine optimal annealing and amplification conditions for the pool of primers (12). Different buffers, salt concentrations, and enzymes have been employed with varying success due to differences in DNA template preparation and the unknown nature of the targeted sequence. We have had the best success by using the following steps in sequence:

1. Prepare eight PCR reaction mixtures on ice, each containing:

   2.5 µl – 10× Gibco PCR buffer
   1.0 µl – $MgCl_2$ (50 mM) – 2 mM final concentration
   2.0 µl – dNTPs (10 mM each)
   0.5 µl – Forward CODEHOP primer (current example is YKQAa primer)
   0.5 µl – Reverse CODEHOP primer (current example is KWDAb primer)
   0.25 µl – Platinum Taq (5 U/µl)
   17.75 µl – Water
   1.0 µl – DNA template (current example is soybean DNA 250 ng/µl)
   25 µl – Total

2. Place the eight tubes in a thermal gradient PCR thermocycler (current example was performed on a Bio-Rad iCycler) such that each tube will have a different annealing temperature ranging from 50 to 65°C.

3. Amplify the DNA using the following conditions:

   1 min – 95°C
   45 cycles of:

   - 30 s – 95°C – melting step
   - 30 s – annealing step (50–65°C gradient)
   - 30 s – 72°C – elongation step

   1 min – 72°C
   Hold – 4°C

4. Analyze amplification products by electrophoresing 5 µl of the PCR reaction on a 2% agarose gel in TAE running buffer. Gels are stained for 5–20 min in ethidium bromide, destained for 5–20 min in distilled water, and visualized under a UV lamp. We use the GelDoc system, which allows us to save pictures of the gels as TIF or JPEG files for presentations or publications. Amplification of soybean DNA using the YKQAa and KWDAWb primers across a 50–65°C annealing gradient is shown in Fig. 5. PCR fragments of ~270 bp were detected at all annealing temperatures except 65°C (lanes 2–9). Optimal amplification was obtained at the lower annealing temperatures

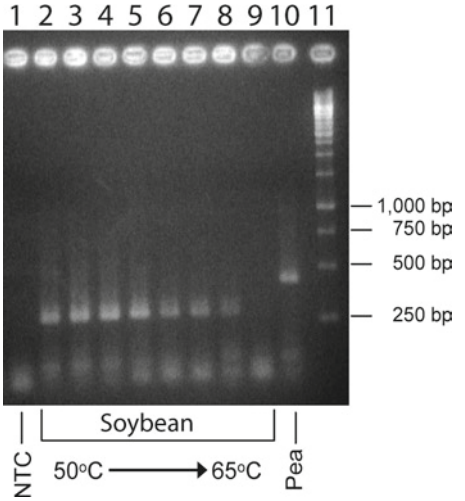


Fig. 5. Thermal gradient PCR amplification of soybean DNA using the YKQAa and KWDAb CODEHOP PCR primers. The YKQAa and KWDAb CODEHOP PCR primers were used to amplify soybean DNA template (250 ng) with a gradient of annealing temperatures from 50–65°C, and the amplification products were analyzed on an ethidium bromide-stained agarose gel. Lane (1) nontemplate control (nonspecific primer/dimer products below 250 bp); Lanes (2–9) YKQAa-KWDAb PCR amplification of soybean DNA template at different annealing temperatures (amplicon size = 269 bp); Lane (10) YKQAa-KWDAb PCR amplification of pea DNA template at 50°C annealing temperature, amplicon size = 416 bp.

(lanes 3, 4). The amplification product obtained from pea DNA (~400 bp) using the same primers is shown in lane 10.

3.5. Sequence Analysis of PCR Product

The PCR product can be gel-purified and directly submitted for sequencing.

1. Take 20 µl PCR reaction and electrophorese on a 2% agarose gel with wide combs.
2. Cut out the band representing the amplicon under a UV lamp with a razor blade. The piece of agarose containing the amplicon DNA is then cut into small pieces and placed into a 1.7-ml microcentrifuge tube.
3. Use a DNA cleanup procedure to extract the DNA out of the agarose. We have good success using the GeneClean III Kit by Q-Biogen, eluting DNA in a final volume of 50 µl buffer. If the band on the gel is very strong, we use up to 200 µl elution buffer.
4. Use 1 µl of purified amplicon DNA to sequence. The sequencing primers consist of a specific primer derived from the consensus clamp region only, of each of the CODEHOP primers. For example, the sequencing primer from the YKQAa CODEHOP is 5′GCTTATTCTCTATGGACTCT, while the sequencing primer from the KWDAWb CODEHOP is 5′CTTCAACAGCCTTCCATGC (see Fig. 3a and b).

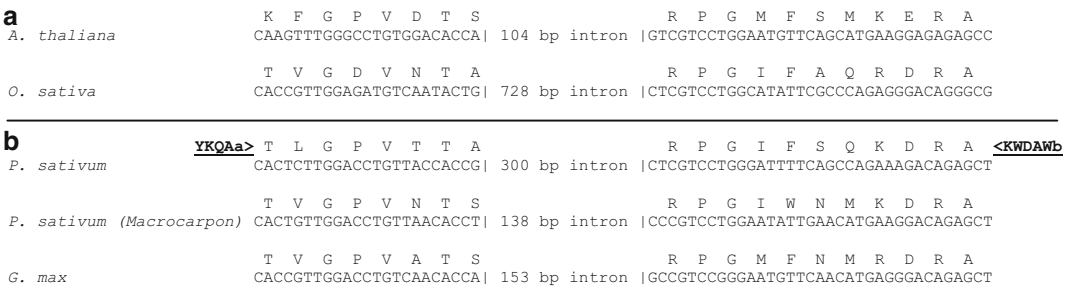


Fig. 6. Alignment of nucleotide and encoded aminoacid sequences of the YKQAa/KWDAb PCR products obtained from pea, snap pea, and soybean with the corresponding sequences of thale cress and rice. The YKQAa/KWDAb CODEHOP PCR products obtained from pea (*Pisum sativum*; 416 bp), snap pea (*Pisum sativum, macrocarpon cultivar*; 254 bp), and soybean (*Glycine max*; 269 bp) were sequenced, and the nucleotide and encoded aminoacid sequences (**b**) were aligned with the corresponding ACBP sequences from thale cress (*A. thaliana*) and rice (*O. sativa*) (**a**) obtained from the NCBI database. The position and size of the intron interrupting the exons encoding the YKQA and KWDA motifs (see Fig. 2) are shown. The genera *Pisum* and *Glycine* both belong to the subfamily *Faboideae* within the family of *Fabaceae*.

We use the ABI version 3.1 sequencing reagents as detailed here:

2 µl Version 3.1 (ABI)
2 µl 2.5× Mix (ABI)
0.5 µl Sequencing primer (use a or b primer)
1 µl PCR product
4.5 µl H_2O
10 µl

This mix is then amplified on a PCR machine (30 cycles of 96°C for 10 s; 50°C for 5 s; 60°C for 4 min) and submitted for sequence determination.

5. Align the resulting novel sequences with each other and with known sequences. When introns are present as in the current example of the ACBP gene, the intron–exon boundaries and intron sizes can be determined from such alignments (Fig. 6). Analysis of the PCR fragments obtained from soybean (G. max) and two species of pea revealed introns of different sizes flanked by two conserved regions of ACBP coding sequences. Thus, three novel ACBP genes (partial sequences) were identified using the YKQAa and KWDAb CODEHOP primer pairs. Using similar approaches, novel genes from different gene families can be readily and quickly identified.

4. Notes

1. For the example presented here, we used the Plant DNA Isolation kit from Qiagen instead of classical proteinase K extraction. However, both methods work well.

2. This mixture was developed for use in the BioRad iCycler, but generic reagents or commercially available master mixes can be used as necessary. In order to obtain the most consistent results and to facilitate reaction setup, it is a good practice to mix all the reagents that are in common. A master mix including the Taq polymerase is stable at –20°C, and aliquots can be stored at 4°C for convenience.
3. For a gene, such as the ACBP gene that contains several introns of diverse lengths, establish the position of conserved exon junctions within the aligned sequences by a TBLASTN search of genomic contigs of the different species that are targeted. This allows the identification of intron positions that could impact the PCR amplification of unknown genes when using genomic DNA as template (see Fig. 2).
4. A DNA fragment amplified between the motifs would be 120–1,500 bp, a size that is easily amplified under normal PCR conditions.
5. Optimal conserved sequence blocks contain 3–4 highly conserved amino acids with restricted codon degeneracy from which the 3′ degenerate core is derived; the presence of serines, arginines, and leucines are not favored due to the presence of six possible codons for each aminoacid. On the other hand, the presence of tryptophan or methionine, each encoded by only one codon, is highly desirable to achieve low degeneracy. In the current example, the C-terminal motif contained seven highly conserved amino acids "RAKWDAW" (Fig. 2); however, the initial arginine and alanine positions were not considered for the degenerate core due to high codon degeneracy.
6. The optimal primer length for a particular CODEHOP PCR primer may vary depending on the DNA template used and should be determined empirically. We typically use CODEHOP PCR primers of 30–32 nucleotides in length.
7. If too many possible CODEHOP primers are displayed, try reducing the degree of degeneracy from the default value of 128 to 64 or 32. If no CODEHOP primers are predicted, examine the phylogenetic analysis of the input protein sequences and remove unrelated or very distantly related sequences from consideration and redo the analysis. In some cases, the input sequences from a targeted gene family may cluster into separate genetic groupings, such as the case with gene paralogs. In this case, primer prediction may be limited to one of the groups to obtain optimal CODEHOP primers.

Acknowledgments

The authors would like to thank Greg Bruce and Elie Karabunarlieva for advice and assistance in preparing this chapter. This work was supported in part by R24 RR021346 from the National Center for Research Resource, NIH and 0833779 from the National Science Foundation.

References

1. Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M., and Henikoff, S. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res* **26**, 1628–35.
2. Rose, T.M., Henikoff, J.G., and Henikoff, S. (2003) CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Res* **31**, 3763–6.
3. Boyce, R., Chilana, P., and Rose, T.M. (2009) iCODEHOP: a new interactive program for designing COnsensus-DEgenerate Hybrid Oligonucleotide primers from multiply aligned protein sequences. *Nucleic Acids Res* **37**, W222–8.
4. Rose, T.M., Schultz, E.R., and Todaro, G.J. (1992) Molecular cloning of the gene for the yeast homolog (ACB) of diazepam binding inhibitor/endozepine/acyl-CoA-binding protein. *Proc Natl Acad Sci U S A* **89**, 11287–91.
5. Gersuk, V.H., Rose, T.M., and Todaro, G.J. (1995) Molecular cloning and chromosomal localization of a pseudogene related to the human acyl-CoA binding protein/diazepam binding inhibitor. *Genomics* **25**, 469–76.
6. VanDevanter, D.R., Warrener, P., Bennett, L., Schultz, E.R., Coulter, S., Garber, R.L., and Rose, T.M. (1996) Detection and analysis of diverse herpes viral species by consensus primer PCR. *J Clin Microbiol* **34**, 1666–71.
7. Rose, T.M., Strand, K.B., Schultz, E.R., Schaefer, G., Rankin, G.W., Jr., Thailess, M.E., Tsai, C.C., and Bosch, M.L. (1997) Identification of two homologs of the Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in retroperitoneal fibromatosis of different macaque species. *J Virol* **71**, 4138–44.
8. CODEHOP: consensus-degenerate hybrid oligonucleotide primers. (1998) (Accessed at http://blocks.fhcrc.org/blocks/codehop.html).
9. Wilson, C.A., Wong, S., Muller, J., Davidson, C.E., Rose, T.M., and Burd, P. (1998) Type C retrovirus released from porcine primary peripheral blood mononuclear cells infects human cells. *J Virol* **72**, 3082–7.
10. Osterhaus, A.D., Pedersen, N., van Amerongen, G., Frankenhuis, M.T., Marthas, M., Reay, E., Rose, T.M., Pamungkas, J., and Bosch, M.L. (1999) Isolation and partial characterization of a lentivirus from talapoin monkeys (*Myopithecus talapoin*). *Virology* **260**, 116–24.
11. Schultz, E.R., Rankin, G.W., Jr., Blanc, M.P., Raden, B.W., Tsai, C.C., and Rose, T.M. (2000) Characterization of two divergent lineages of macaque rhadinoviruses related to Kaposi's sarcoma-associated herpesvirus. *J Virol* **74**, 4919–28.
12. Rose, T.M., Ryan, J.T., Schultz, E.R., Raden, B.W., and Tsai, C.C. (2003) Analysis of 4.3 kilobases of divergent locus B of macaque retroperitoneal fibromatosis-associated herpesvirus reveals a close similarity in gene sequence and genome organization to Kaposi's sarcoma-associated herpesvirus. *J Virol* **77**, 5084–97.
13. Rose, T.M. (2005) CODEHOP-mediated PCR – a powerful technique for the identification and characterization of viral genomes. *Virol J* **2**, 20.
14. Staheli, J.P., Ryan, J.T., Bruce, A.G., Boyce, R., and Rose, T.M. (2009) Consensus-degenerate hybrid oligonucleotide primers (CODEHOPs) for the detection of novel viruses in non-human primates. *Methods* **49**, 32–41.
15. Bruce, A.G., Bakke, A.M., Gravett, C.A., DeMaster, L.K., Bielefeldt-Ohmann, H., Burnside, K.L., and Rose, T.M. (2009) The ORF59 DNA polymerase processivity factor homologs of Old World primate RV2 rhadinoviruses are highly conserved nuclear antigens expressed in differentiated epithelium in infected macaques. *Virol J* **6**, 205.

16. Burton, M., Rose, T.M., Faergeman, N.J., and Knudsen, J. (2005) Evolution of the acyl-CoA binding protein (ACBP). *Biochem J* **392**, 299–307.
17. Bio-itest. Northwest Association for Biomedical Research. (Accessed at http://www.nwabr.org/education/itest.html).
18. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–8.
19. iCODEHOP: Interactive Program for Creating Consensus-Degenerate Hybrid Oligonucleotide Primers. (2008) (Accessed at https://icodehop.cphi.washington.edu/i-codehop-context/Welcome).

Chapter 6

Sequencing of Difficult DNA Regions by SAM Sequencing

Keith R. Mitchelson

Abstract

Nucleotide analogues are used increasingly in medicine and biotechnology to effect DNA sequence change, principally via clastogenic and transcriptional effects. This article, however, discusses the use of mutagenic nucleotide analogues to improve the sequencing of recalcitrant and repetitive DNA motifs. Guidance in the technical and practical approaches that support use of this approach with different DNA sequencing technologies is provided, including for high-throughput technologies.

Key words: Mutagenic nucleotide analogue, DNA sequencing

1. Introduction

Several new ultrahigh-throughput sequencing technologies have become recently available which complement, and to a large extent, supersede standard Sanger sequencing (1, 2), yet those technologies that require steps involving use of PCR or sequence extension polymerases may continue to encounter regions of sequence that are problematic. Identification of the error in the determined sequence of such regions may not be readily recognized and often even when recognized cannot be readily resolved.

A technique termed SAM sequencing (Sequencing Aided by Mutagenesis), which employs advanced reconstruction algorithms to extract the correct target sequence from several randomly mutated copies of a DNA target, has been demonstrated both theoretically (3–6) and practically for small recalcitrant target regions (7, 8).

The introduction of random replacement mutations does not destroy original sequence information, but instead distributes it amongst multiple variants. These variants may lack many of the problematic features of the initial target and are then much more

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_6, © Springer Science+Business Media, LLC 2011

amenable to sequencing. Because the approach involves changing the target sequence, it can address difficulties arising from any problematic sequence characteristic and relieve confusion due to repeated motifs. SAM techniques can also simplify the assembly of repeated regions by introducing "landmarks" that distinguish between different repeats. The introduction of landmark mutations also benefits sequence reconstruction from multiple, short overlapping reads into longer regions, necessary for de novo sequencing. Currently, mutagenic nucleotide analogues have demonstrated improved sequencing of small intractable DNA regions (7), yet their potential for improving sequencing of significantly larger genomic regions has not been fully developed.

1.1. Mutation Using Nucleotide Analogues

Mutagenic nucleoside analogue moieties are compounds that may mimic natural nucleoside in structural associations with natural nucleosides in DNA, permitting specific bases to be targeted by the choice of analogue used. Ideally, a mutagenic analogue for SAM sequencing would induce changes at completely random sites. Practically, however, mutagenic analogues do not induce mutations completely at random rather they induce preferred nucleotide transition reactions (with rare transversions, insertions, and deletions) in an almost random distribution (9–14). The different random mutations cause individual mutant copies to have different sequences which may be separated (cloned) and individually sequenced. Mutations introduced at very early rounds of mutagenic-PCR amplification can establish "founder mutations" that occur in a significant proportion of the progeny amplimers, although these founder mutations are themselves at random loci (Fig. 1). This characteristic allows the influence of founder mutations to be minimized through simple experimental design. The DNA sequences determined from a low number of the altered copies can then be analyzed using Bayesian methods (3, 5) to reconstruct the original wild-type sequence. Remarkably, the entire process has efficiencies and accuracies roughly equivalent to those of conventional sequencing (15).

1.2. In Vitro Genomic Mutation

We have previously described a workflow involving the PCR mutation amplification of targets in the presence of mutagenic analogues, then cloning the amplified fragment and sequencing of multiple copies. This approach is not very efficient; hence, we sought methods to multiplex the mutation of numerous potential targets, which then require only the recovery of the target region(s) for subsequent sequencing.

Fig. 1. (continued) 200 μM BrdUTP during amplification. The inclusion of dPTP counteracts the effect of BrdUTP, by inducing A to G and T to C mutations, which increase the GC content. The effect of the two mutagens is a more random and balanced distribution of mutation along the DNA sequence, while maintaining the overall mutation level at around 5.5%. Reproduced from (15) with permission. Copyright 2007 Elsevier Science.

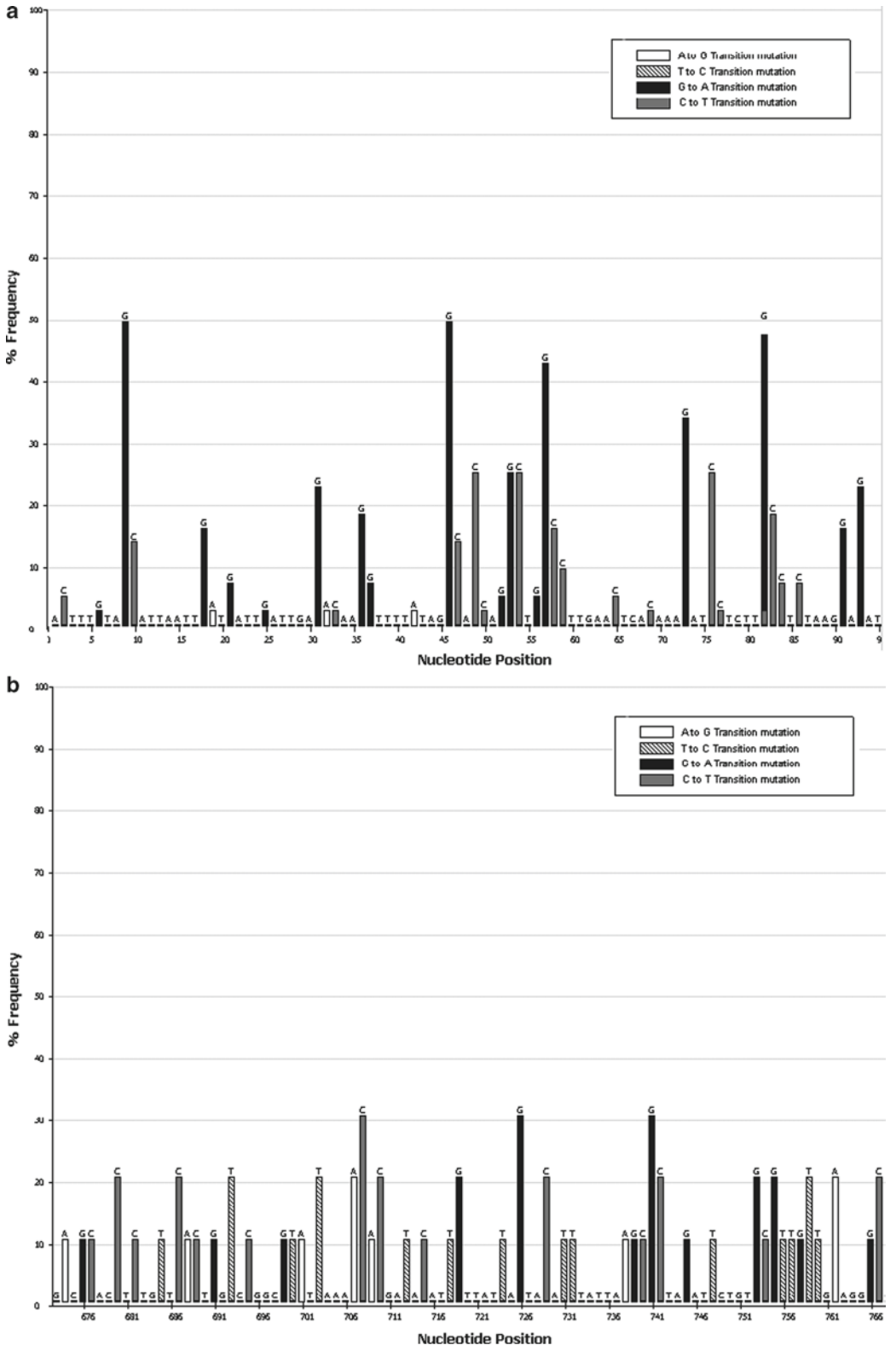


Fig. 1. PCR mutation frequencies observed with combinations of mutagenic analogues. (**a**) PCR mutation frequencies observed for 45 sequences exposed to 400 μM BrdUTP alone. The BrdUTP introduces G to A and C to T mutations that reduce the overall GC content. (**b**) PCR mutation frequencies detected in four sequences exposed to 200 μM dPTP and

A rolling circle mechanism (RCA) method based on the ϕ29 DNA polymerase (16–18) has been widely used for in vitro amplification of plasmid DNA for sequencing directly from single bacterial colonies. It has also been used to enable sequencing of low-copy plasmids, which give poor yields from mini-preparations. We employed RCA for the introduction of random mutations in large DNA molecules which can then template for conventional PCR amplification of smaller discrete target regions which are then cloned and sequenced. The ability to amplify large DNAs in vitro offers the opportunity of directing mutation to higher levels, simultaneously mutating numerous PCR-able regions. The RCA method also avoids the functional limits imposed by replication in bacterial hosts.

1.3. Megabase Regions

We have recently reported a theoretical reconstruction analysis which suggested SAM methods would be an efficient and cost-effective aid to high-throughput de novo sequencing of megabase-sized DNA fragments and small genomes (15). Although we have not applied SAM practically to any of the new ultrahigh-throughput sequencing methods, our theoretical analysis suggests that it should be feasible. With many of the high-throughput, new-generation DNA sequencing (NGS) technologies (1), the amplification and capture of multiple copies of short target regions are intrinsic to the method; hence, many of the wet-lab procedures described here as individual steps could be avoided, and the integration of SAM mutation techniques could be exploited to benefit the reading of repeated sequences, homopolymer elements, and multimer nucleotide repeat tracts, for example, AT-rich repeat elements. Further development of sequence alignment algorithms would be needed for use of SAM methodologies with short-read sequencing technologies, principally to aid identification and alignment of mutated copies of the same target sequence from a pool of genomic sequence reads.

Many of the new short-read sequencing technologies suffer from an inability to traverse repeat sequence regions and are subject to systematic bias and remarkable inefficiencies in the reading of particular motifs. Harismendy and colleagues (19) compared the same 260 kb human sequence from four individuals generated by three NGS platforms, Roche 454, Illumina GA, and ABI SOLiD, which overlapped with 88 kb of ABI 3730xL Sanger sequence. At high coverage depth, they observed that the base calling errors with NGS platforms are systematic and occur as a result of miscalling within local sequence contexts, and as the depth of coverage is decreased, additional "random sampling" errors in base calling occur. Each NGS platform also generated its own patterns of biased sequence coverage. For the short-read platforms (Illumina GA, ABI SOLiD), AT-rich repetitive sequence was poorly covered, and the ends of fragments were massively

overrepresented. Dohm and colleagues (20) also showed that the Illumina 1G platform substantially overrepresented sequence with elevated GC content and underrepresented elevated AT content regions, as well as showing increased miscalling of bases preceded by a G. Overall, approximately 50% of the output from the short-read platforms was not usable, although some ~95% of the Roche 454 sequence could be aligned uniquely to the target sequence (19, 20) despite its known inability to read A/T homopolymer tracts accurately (21). Homopolymer tracts and AT-rich regions are two classes of repetitive sequence motif for which SAM sequencing has demonstrated benefit.

2. Materials

2.1. Reagents

1. Standard nucleotide triphosphates (dTTP, dATP, dGTP, dCTP) and desalted oligonucleotide primers (Roche Applied Science).
2. Universal M13 oligonucleotide primers: for -20 (Cat. #S121s) and rev -24 (Cat. #S1201s); for -40 (Cat. #S1212s) and rev -48 (Cat. #S1233s) (New England Biolabs).
3. High efficiency Ampli*Taq* Gold polymerase and Sanger BigDye v 2.0 and v 3.0 cycle sequencing kits (Applied Biosystems).
4. TempliPhi ϕ29 DNA polymerase (Amersham Biosciences).
5. Mutagenic nucleotide triphosphate (mNTP) analogues were purchased from Trilink BioTechnologies and Sigma-Aldrich.
6. Cloning vectors were pGEM T-EASY (Promega) and pDrive (Qiagen) and plasmids were cloned in competent *E. coli* JM109 cells.
7. Plasmids were isolated from single clones using REAL Prep (Qiagen).

2.2. Mutagenic Nucleotide Triphosphates

The most useful mutagenic nucleotide analogues which are capable of being efficiently incorporated into DNA include 6-(2-deoxy-B-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C] oxazin-7-one triphosphate (dPTP or dQ_6TP), bromodeoxyuridine triphosphate (BrdUTP), 8-oxo-2′-deoxyadenosine (OdATP), and 8-oxo-2′-deoxyguanosine (OdGTP, d8-oxo-GTP). These nucleotides are obtainable from TriLink BioTechnologies.

Nucleotide triphosphates such as 5-hydroxy-deoxycytosine (5OH-dCTP), and triphosphates of more unusual nucleotides such as 5-fluoro-2′-deoxyuridine (FldU), 5-formyl-2′-deoxyuridine (fdU), 4-methyl, 5-hydroxy-deoxycytosine (4Me 5OH-dCTP, methoxy-dC), N^6-methoxy-2,6-diaminopurine

(dKTP), and N^6-methoxyadenine (dZTP) are incorporated less well or poorly but add to the repertoire of potentially useful analogues. Some are obtainable from TriLink BioTechnologies as stock, or as custom-synthesized products.

3. Methods

3.1. PCR Amplification Mutation of DNA Fragments

3.1.1. Mutation Reactions

Standard PCR amplifications use 2 ng of DNA template, 1× AmpliTaq Gold buffer (ABI), 200 μM dNTPs, 2 mM $MgCl_2$, 0.4 μM each primer, and 1 U of Ampli*Taq* Gold polymerase (ABI) in a total of 25 μL. Mutation reactions are essentially as described by Zaccolo and colleagues (9) in which PCR mixes were supplemented individually with 200–400 μM of the triphosphate esters of mutagenic analogues (9–14) dPTP, OdGTP, OdATP, 5-BrdUTP, dZTP, or dKTP and with up to 400 μM of additional $MgCl_2$. Amplifications are typically 30 cycles, with 60 s and 30 s periods of denaturation and annealing respectively, and extension for periods up to 10 min (see Note 1).

3.1.2. Cloning Methods

The target sequence used here for illustration was an undefined DNA fragment of 1.5 kb in length cloned into pUC19. Universal M13 primers (FSP-21, FSP-40, RSP-26, RSP-48) were used for PCR amplification. Using a concentration of 400 μM of each dATP, dCTP, dGTP, dTTP, and additional single mutagenic nucleotides such as either dPTP or BrdUTP or 8-oxo-GTP (at 400 μM) or combinations of two appropriate mutagenic nucleotides (e.g., 200 μM dPTP and 200 μM BrdUTP) in a standard PCR. Mutations were incorporated up to a frequency of 1 in 5. The action of the analogues can be investigated individually, or two analogues together (see Fig. 1).

3.2. Varying the Mutation Intensity

Low and high mutation intensities were also established with 400 μM dPTP by reducing or increasing the cycling number between 20 and 38 times, respectively; with additional recovery/amplification PCR steps with standard dNTP mixes for the low intensity mutations at reduced cycle number (see Note 2).

3.2.1. Reaction Conditions

2 ng DNA template, 1× AmpliTaq Gold buffer, 400 μM dNTPs, 2 mM magnesium chloride, 0.4 μM each primer, and 1 U of AmpliTaq Gold, in a total of 25 μL. The cycling reactions are performed as follows: 1 cycle of 94°C for 15 min, 30 cycles of 94°C for 1 min, 50°C for 0.5 min, 72°C for 5 min, 1 cycle of 72°C for 10 min. This regimen yields PCR products incorporating analogue bases. Clonable mutated DNA fragments are recovered by reamplification of 1 μL of the reaction products with nested primers and the four natural dNTPs and standard PCR conditions.

3.3. Cloning of PCR Products and DNA Sequencing

3.3.1. Cloning of Mutagenized PCR Products

The products of the PCR recovery can be ligated into pGEM T-EASY or pDrive vectors and transformed into competent *E. coli* JM109 cells. Plasmids can be isolated from single clones using REAL Prep (Qiagen).

3.3.2. Clone Sequencing

In the illustrated studies, individual clones were cycle-sequenced with Applied BioSystems BigDye v2.0 and v3.0 sequencing kits using the manufacturer-recommended conditions. The M13 oligonucleotide primers for -20, rev -24, for -40, and rev -48 Cat. (New England Biolabs) were used for PCR amplification and for cycle sequencing.

3.3.3. Sequence Analysis

The sequence data from individual clones was analyzed for mutation frequency, and collectively the data from several clones may be analyzed using SAM algorithms (4–6) to enable the reconstruction of the original sequence.

3.4. Minimizing Founder Mutations

Mutations introduced at very early rounds of mutagenic-PCR amplification can establish "founder mutations" that occur in a significant proportion of the progeny amplimers and which may confound extraction of the correct original sequence. As these founder mutations are themselves typically established at random loci at low intensities of mutation, experimental design can minimize their influence.

3.4.1. Use of Separate Mutagens

Several separate mutation PCR amplification reactions can be undertaken and sequence determined from sets of clones from each mutation process. These reactions can involve the use of two different mutagenic nucleotides to tend to establish different founder mutations. Often the DNA target is rich in a particular base which is the original cause of the difficulty to sequence. Modification of these homopolymer tracts or base-biased sequence elements can only be achieved by appropriate mutagens. Hence, it is advantageous to having a choice of mutagenic analogues effecting the same transversion or transition reactions, albeit at different efficiencies (see Table 1).

3.4.2. Use of Pairs of Mutagens

PCR mutation amplification of template with equimolar pairs of mutagenic nucleotides, for example, 200 μM dPTP and 200 μM BrdUTP, introduces a broader distribution of mutation events along the target molecule, each analogue effecting mutation of different target bases. The BrdUTP introduces G to A and C to T mutations that effectively reduce the GC content. The inclusion of dPTP counteracts the overall effect of BrdUTP by inducing A to G and T to C mutations and increasing the GC content. The effect of the two mutagens is a more random distribution of

Table 1
Predominant mutation events affected by DNA amplification/replication in the presence of particular nucleotide analogues (see Note 5)

Mutagenic nucleotide	1° Event	2° Event
dPTP[a]	A→G transitions	T→C transitions
dKTP	A→G transitions	T→C transitions
dZTP	A→G transitions	T→C transitions
8-oxo-dGTP[b]	A→C transversions	T→G transversions
5-bromo-2′-deoxyuridine (BrdU)[a]	T→C transitions	A→G transitions
5-fluoro-2′-deoxyuridine (FldU)	T→C transitions	A→G transitions
5-formyl-2′-deoxyuridine (fdU)	T→C transitions	A→G transitions
5OH-dCTP[a]	G→A transitions	C→T transitions
8-oxo-dATP[b]	G→A transitions	C→T transitions
5-methyl-N4-hydroxydeoxycytidine	C→T transitions	G→A transitions
2-aminopurine	T→C transitions	A→G transitions

[a]Efficiently incorporated mNTP
[b]Moderately incorporated mNTP

mutation and a more balanced distribution of mutations along the DNA sequence, yet maintaining the overall mutation level at a low level of around 5.5% (Fig. 1). This approach can be used generally for many sequences, but has some limited use with strongly base-biased targets or recalcitrant homopolymer regions (see Note 2).

3.5. Rolling Circle Amplification

We have incorporated nucleotide analogues such as dPTP and/or BrdUTP directly into DNA templates by RCA. The target region is then recovered by standard PCR. Cloning and sequencing of copies of target regions provided sequence variants from which the correct original sequence is reconstructed after SAM analysis.

3.5.1. Reaction Conditions

Amplification can be effected from circular DNA templates using either purified DNA or bacterial cultures. The final Phi29 reaction buffer can contain 40 mM Tris–HCl (pH 7.5), 50 mM KCl, 10 mM $MgCl_2$, 5 mM $(NH_4)_2SO_4$, and 4 mM DTT. Either purified circular DNA template or a single colony pick can be used to template the reaction.

3.5.1.1. Dilution of the Template Colony

Take only a small part of one colony on the tip of a sterile toothpick. In order to get the greatest level of amplification, as little of the colony as possible should be taken to limit the amount of contamination (cell debris, agar, etc.). Add the colony pick to 100 µL of water in a PCR tube and mix briefly to dislodge some cells. Serially dilute the cell number by transferring 1 µL from this mixture into 4 µL of water in a second PCR tube. Then, dilute a further 1 µL of this dilution into a third PCR tube containing 4 µL of water. Use the third dilution for amplification (see Note 3).

3.5.1.2. Purified DNA

Typically, use 2 ng purified circular BAC DNA template. Use from 1 pg to 10 ng of DNA template with increasing amounts as template size increases. The amplification reaction is carried out in 1× Phi29 reaction buffer, 400 µM dNTPs, 200–400 µM nucleotide analog(s), 2 mM magnesium chloride (additional), 0.4 µM primer, and 1 U of RepliPhi29 DNA Polymerase in a total of 20 µL. Nucleotides are at concentrations of 400 µM for each dATP, dCTP, dGTP, dTTP, plus additional single mutagenic nucleotides such as either dPTP or BrdUTP or 8-oxo-GTP at concentrations from 100 to 400 µM.

3.6. Mutagenic TempliPhi Amplification of purified Plasmids or BACs

Use from 1 pg to 10 ng of DNA template (increasing amounts are used as template length increases), l× denaturing buffer with 0.4 µM random hexamer primers, l× TempliPhi DNA amplification buffer, 200 µM dNTPs, 50–100 µM analogue dNTP, 2 mM magnesium chloride, and 1 U of Phi 29 DNA polymerase in a total of 20 µL.

Reactions are performed as follows: Mix template DNA and denaturing buffer with random hexamer primers in a volume of 10 µL and denature with 1 cycle of 95°C for 3 min, cool to room temperature or 4°C. Add 1–2 µL 1 mM analogue dNTP, 10 µL of TempliPhi premix (with dNTPs, TempliPhi incubation buffer, and Phi29 DNA polymerase) to the cooled sample and mix briefly, then incubate at 30°C for 8–12 h (recommended range 4–18 h). Heat-inactivate the enzyme at 65°C for 10 min, then cool to 4°C.

3.7. Alkaline Mutagenic TempliPhi Reaction

Use a concentration of 200 µM of each dATP, dCTP, dGTP, dTTP, 50–100 µM analogue dNTP such as 5′-Bromo-dUTP in a mutagenic TempliPhi reaction modified by addition of Glycine KOH, pH 9.5 to a final pH of 8.5–9.0, 2 mM magnesium chloride, and 1 U of Phi29 DNA polymerase in a total of 20 µL (see Note 4).

3.8. Sequencing

Dilute the amplified DNA approximately fivefold and use 10 ng DNA (typically 2–3 µL) as template for a standard DNA sequencing reaction or PCR recovery reaction prior to fragment cloning.

3.8.1. Sequencing of Mutagenized Clones

At higher levels of mutation, individual clones of mutagenized DNA templates need to be recovered prior to sequencing.

Clonable mutated DNA fragments can be recovered by reamplification of 1 μL of the reaction products with nested primers and the four natural dNTPs and standard PCR conditions. The PCR products can be gel-purified and cloned into the pGEM T-EASY vector (Promega) and transformed into *E. coli*. Plasmids DNA from individual clones can be sequenced as above and subjected to SAM reconstruction analysis.

3.8.2. Direct Sequencing

Since the mutagenic analogues are introduced essentially at random, at sufficiently low levels of mutation, the disruption of barriers to DNA polymerase processing may be achieved and the mixture of different mutagenized templates can be directly Sanger-sequenced without impairing the sequencing readout. The level of mutagenesis that may be used successfully for direct sequencing must be empirically determined. Dilute the amplified DNA approximately fivefold and use 10 ng DNA (typically 2–3 μL) as template directly for a standard Sanger sequencing reaction.

4. Notes

1. The analogues are incorporated progressively during PCR at a relatively low frequency per cycle, accumulating mutations with increasing numbers of cycles (9, 11, 12) and resulting in randomly mutated DNA fragments. Although intensities up to 5–10% mutation can be achieved by several mutagenic analogues (10–12), few reach the 20–30% mutation intensities achieved by dPTP (9, 22). Mutagenic analogues tautomerize to establish equilibrium between amino and imino forms (10, 13), and the ratio of this equilibrium and the differing potential of the forms to be incorporated by polymerase during replication result in a random pattern of mutation within a given DNA sequence. The different tautomeric forms replace several different natural nucleosides rather than a single *cognate* nucleoside. This induces transition and transversion mutations in subsequent rounds of DNA replication when a novel cognate nucleoside, initially introduced opposite the nucleotide analogue, itself forms base pairs with its natural cognate.
2. The choice of mutagenic analogue depends upon base content of the problem sequence and the intensity of mutation that is required to remove the barrier to sequencing or cloning. Manipulation of the DNA amplification conditions and the nucleoside analogue concentration also readily affords control over the intensity of mutation achieved when using high-intensity mutators (9).

3. The cotransformation of repair-deficient bacteria with plasmid and analogues, or extended growth of bacteria carrying plasmids, phage, and BACs in the presence of analogues is not efficient, as damage to hosts can severely limit the recovery of clones (23, 24).
4. The ratio between amino and imino forms can be influenced by pH, as can the efficiency of incorporation of both nucleosides and analogues by DNA polymerases – providing for additional control in the direction of mutagenic transversions and transitions (10, 13). The incorporation efficiencies of 5-bromodeoxyuridine and 5-fluorodeoxyuridine opposite guanosine increase by about 20× and 13× respectively as the reaction pH is raised from 7.0 to 9.5 (10). In contrast, other mutagenic analogues are believed to act through transitional structural forms that are not influenced by ionization (22).
5. Fuller details for a number of nucleotide-like analogues which are capable of introducing mutations into DNA are: (1) dPTP (or dQ_6TP) [6-(2-deoxy-B-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C]oxazin-7-one triphosphate] behaves as thymine in the majority of DNA copying events and as cytosine in a minority, inducing A to G and T to C transitions, which raises the GC content (and lowers the AT content) (9, 26). (2) 8-oxo-dGTP preferentially causes A to C and T to G transversions (9, 25) although the tautomerization constant and attendant frequency of mutation is low. (3) N^6-methoxy-2,6-diaminopurine (dK) behaves as adenine in the majority DNA copying events and as guanine in a minority, preferentially causing A to G and T to C transitions (11, 26). (4) N^6-methoxyadenine (dZ) behaves as adenine in the majority of DNA copying events and as guanine in a minority, preferentially causing A to G and T to C transitions (11, 26). (5)5-bromo-2′-deoxyuridine(BrdU),5-fluoro-2′-deoxyuridine (FldU), and 5-formyl-2′-deoxyuridine (fdU) each behave as thymine in the majority of DNA copying events and as cytosine in a minority, principally inducing G to A and C to T mutations, which increases the overall AT content (10, 14). These mutations may be influenced by the ionization state of the analogue and are enhanced at elevated pH values. The ratio of transitions to transversions is altered if DNA amplification is performed at elevated pH. (10, 27–29). (6) 8-oxo-dATP induces GC to AT transitions, which increases the overall AT content. The ratio of transitions to transversions is altered if DNA amplification is performed at elevated pH as a result of the formation of particular tautomeric forms at elevated pH (10, 27, 28). (7) N^4-methoxycytosine induces purine transition mutagenesis. When methoxylated dCTP is incorporated into DNA, it behaves as thymine in the majority DNA copying events and as cytosine in a minority. However,

once incorporated in the template strand, methoxycytosine directs the incorporation of adenine above guanosine in the majority of cases (30). (8) The mutagenic specificity of 5-methyl-N4-hydroxydeoxycytidine appears to be opposite to that of 2-aminopurine. This suggests that the former can induce transition of CG to TA, while the latter TA to CG (31). (9) The analogues 5-hydroxy-deoxycytosine (5OH-dCTP) and 4-methyl, 5-hydroxy-deoxycytosine (4Me 5OH-dCTP) can replace dCTP, and to a lesser extent dTTP. Once incorporated, the analogues can template for particular nucleotides. dG is predominantly incorporated opposite 5-OH-dC, with low dA incorporation also seen. 5OH-dCTP has the principal mutagenic potential for G to A transitions (32–34). (10) Other potentially useful mutagenic nucleotides have been reported but are not currently commercially available. These include compounds such as a 2-chloro-4-fluoroindole nucleoside (an isosteric analogue of 8-oxo-dG) and a 2-chloro-4-methylbenzimidazole nucleoside (a mimic of 8-oxo-dA) (35).

References

1. Voelkerding, K.V, Dames, S.A., and Durtschi, J.D. (2009) Next-generation sequencing: from basic research to diagnostics. *Clin. Chem.* **55**, 641–58.
2. Mitchelson, K.R., Hawkes, D.B., Turakulov, R., and Men, A.E. (2007) Overview: developments in DNA sequencing. In: *New High Throughput Technologies for DNA Sequencing and Genomics.* Series: *Perspectives in Bioanalysis* (Ed., Mitchelson, K.R.), vol. **2**, Elsevier, Amsterdam, pp. 3–44.
3. Keith, J.M., Adams, P., Bryant, D., Mitchelson, K.R., Cochran, D.A.E., and Lala, G.H. (2003) Inferring an original sequence from erroneous copies: a Bayesian approach. In: *Proceedings of the 1st Asia-Pacific Bioinformatics Conference* (*APBC2003*) (Ed., Chen, Y.P.), vol. **19**, Australian Computer Society, Adelaide, pp. 23–28.
4. Keith, J.M., Adams, P., Bryant, D., Mitchelson, K.R. Cochran, D.A.E., and Lala, G.H. (2003) Inferring an original sequence from erroneous copies: two approaches. *Asia Pacific Biotech News* 7, 107–14.
5. Keith, J.M., Adams, P., Bryant, D., Cochran, D.A.E., Lala, G.H., and Mitchelson, K.R. (2004) Algorithms for sequencing via mutagenesis. *Bioinformatics* **20**, 2401–10.
6. Keith, J.M., Adams, P., Bryant, D., Kroese, D.P., Mitchelson, K.R., Cochran, D.A.E., and Lala, G.L. (2002) A simulated annealing algorithm for finding consensus sequence. *Bioinformatics* **18**, 1494–9.
7. Keith, J.M., Cochran, D.A.E., Lala, G.H., Adams, P., Bryant, D., and Mitchelson, K.R. (2004) Unlocking hidden genomic sequence. *Nucleic Acids Res.* **32**, e35.
8. Cochran, D., Lala, G., Keith, J., Adams, P., Bryant, D., and Mitchelson, K. (2006) Sequencing by aligning mutated DNA fragments. In: *The Frontiers of Biochip Technologies* (Eds., Xing, W.L. and Cheng, J.), Springer, New York, pp. 231–45.
9. Zaccolo, M., Williams, D.M., Brown, D.M., and Gherardi, E. (1996) An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* **255**, 589–603.
10. Yu, H., Eritja, R., Bloom, L.B., and Goodman, M.F. (1993) Ionization of bromouracil and fluorouracil stimulates base mispairing frequencies with guanine. *J. Biol. Chem.* **268**, 15, 935–43.
11. Hill, F., Williams, D.M., Loakes, D., and Brown, D.M. (1998) Comparative mutagenicities of N6-methoxy-2,6-diaminopurine (dK) and N6-methoxyaminopurine 2 -deoxyribonucleosides (dZ) and their 5 -triphosphates. *Nucleic Acids Res.* **26**, 1144–9.
12. Kamiya, H., Maki, H., and Kasai, H. (2000) Two DNA polymerases of *Escherichia coli* display distinct misinsertion specificities for 2-hydroxy-dATP during DNA synthesis. *Biochemistry* **39**, 9508–13.
13. Suen, W., Spiro, T.G., Sowers, L.C., and Fresco, J.R. (1999) Identification by UV

resonance Raman spectroscopy of an imino tautomer of 5-hydroxy-2′-deoxycytidine, a powerful base analog transition mutagen with a much higher unfavored tautomer frequency than that of the natural residue 2′-deoxycytidine. *Proc. Natl. Acad. Sci. USA* **96**, 4500–5.

14. Ma, X., Ke, T., Mao, P., Jin, X., Ma, L., and He, G. (2008) The mutagenic properties of BrdUTP in a random mutagenesis process. *Mol. Biol. Rep.* **35**, 663–7.
15. Keith, J.M., Hawkes, D.B., Carter, J.A., Cochran, D.A.E, Adams, P., Bryant, D.E., and Mitchelson, K.R. (2007) Sequencing aided by mutagenesis facilitates *de novo* sequencing of megabase DNA fragments by short length reads. In: *New High Throughput Technologies for DNA Sequencing and Genomics.* Series: *Perspectives in Bioanalysis* (Ed., Mitchelson, K.R.), vol. **2**, Elsevier, Amsterdam, pp. 303–26.
16. Dean, F.B., Nelson, J.R., Giesler, T.L., and Lasken, R.S. (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* **11**, 1095–9.
17. Nelson, J.R., Cai, Y.C., Giesler, T.L., Farchaus, J.W., Sundaram, S.T., Ortiz-Rivera, M., Hosta, L.P., Hewitt, P.L., Mamone, J.A., Palaniappan, C., and Fuller, C.W. (2002) TempliPhi, ϕ29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. *Biotechniques* **32**, S44–7.
18. Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S.F., Egholm, M., and Lasken, R.S. (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA* **99**, 5261–6.
19. Harismendy, O., Ng, P.C., Strausberg, R.L., Wang, X., Stockwell, T.B., Beeson, K.Y., Schork, N.J., Murray, S.S., Topol, E.J., Levy, S., and Frazer, K.A. (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* **10**, R32.
20. Dohm, J.C., Lottaz, C., Borodina, T., and Himmelbauer, H. (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* **36**, e105.
21. Wicker, T., Schlagenhauf, E., Graner, A., Close, T.J., Keller, B., and Stein, N. (2006) 454 sequencing put to the test using the complex genome of barley. *BMC Genomics* **7**, 275.
22. Schuerman, G.S., Van Meervelt, L., Loakes, D., Brown, D.M., Kong Thoo Lin, P., Moore, M.H., and Salisbury, S.A. (1998) A thymine-like base analogue forms wobble pairs with adenine in a Z-DNA duplex. *J. Mol. Biol.* **282**, 1005–11.
23. Inoue, M., Kamiya, H., Fujikawa, K., Ootsuyama, Y., Murata-Kamiya, N., Osaki, T., Yasumoto, K., and Kasai, H. (1998) Induction of chromosomal gene mutations in *Escherichia coli* by direct incorporation of oxidatively damaged nucleotides. *J. Biol. Chem.* **273**, 11069–74.
24. Negishi, K., Loakes, D., and Schaaper, R.M. (2002) Saturation of DNA mismatch repair and error catastrophe by a base analogue in *Escherichia coli*. *Genetics* **161**, 1363–71.
25. Nampalli, S. and Kumar, S. (2000) Efficient synthesis of 8-oxo-dGTP: a mutagenic nucleotide. *Bioorg. Med. Chem. Lett.* **10**, 1677–9.
26. Hill, F., Loakes, D., and Brown, D.M. (1998) Polymerase recognition of synthetic oligodeoxyribonucleotides incorporating degenerate pyrimidine and purine bases. *Proc. Natl. Acad. Sci. USA* **95**, 4258–63.
27. Driggers, P.H. and Beattie, K.L. (1988) Effect of pH on the base-mispairing properties of 5-bromouracil during DNA synthesis. *Biochemistry* **27**, 1729–35.
28. Sowers, L.C., Fazakerley, G.V., Eritja, R., Kaplan, B.E., and Goodman, M.F. (1986) Base pairing and mutagenesis: observation of a protonated base pair between 2-aminopurine and cytosine in an oligonucleotide by proton NMR. *Proc. Natl. Acad. Sci. USA* **83**, 5434–8.
29. Privat, E.J. and Sowers, L.C. (1996) A proposed mechanism for the mutagenicity of 5-formyluracil. *Mutat. Res.* **354**, 151–6.
30. Hossain, M.T., Sunami, T., Tsunoda, M., Hikima, T., Chatake, T., Ueno, Y., Matsuda, A., and Takénaka, A. (2001) Crystallographic studies on damaged DNAs IV. N(4)-methoxycytosine shows a second face for Watson-Crick base-pairing, leading to purine transition mutagenesis. *Nucleic Acids Res.* **29**, 3949–54.
31. Janion, C. (1978) The efficiency and extent of mutagenic activity of some new mutagens of base-analogue type. *Mutat. Res.* **56**, 225–34.
32. Purmal, A.A., Kow, Y.W., and Wallace, S.S. (1994) 5-Hydroxypyrimidine deoxynucleoside triphosphates are more efficiently incorporated into DNA by exonuclease-free Klenow fragment than 8-oxopurine deoxynucleoside triphosphates. *Nucleic Acids Res.* **22**, 3930–5.

33. Purmal, A.A., Kow, Y.W., and Wallace, S.S. (1994) Major oxidative products of cytosine, 5-hydroxycytosine and 5-hydroxyuracil, exhibit sequence dependent mispairing *in vivo*. *Nucleic Acids Res.* **22**, 72–8.
34. Loeb, L.A., Essigmann, J.M., Kazazi, F., Zhang, J., Rose, K.D., and Mullins, J.I. (1999) Lethal mutagenesis of HIV with mutagenic nucleoside analogs. *Proc. Natl. Acad. Sci. USA* **96**, 1492–7.
35. Taniguchi, Y. and Kool, E.T. (2007) Nonpolar isosteres of damaged DNA bases: effective mimicry of mutagenic properties of 8-oxopurines. *J. Am. Chem. Soc.* **129**, 8836–44.

Chapter 7

A Global Single-Cell cDNA Amplification Method for Quantitative Microarray Analysis

Kazuki Kurimoto and Mitinori Saitou

Abstract

We describe here a protocol to faithfully amplify global cDNAs from single cells. The amplified cDNAs retain their sense–antisense orientation and can be easily applied to template preparation for quantitative high-density oligonucleotide microarray analyses. The amplification protocol comprises (1) lysis of a single cell in a tube without purification, (2) first-strand cDNA synthesis with the first primer tailed with oligo dT, (3) elimination of the unreacted first primer, (4) poly (dA) tailing of the cDNA, (5) second-strand cDNA synthesis with the second primer tailed with oligo dT, and (6) 20-cycle, directional PCR with the two primers. To prepare the template for the isothermal linear amplification with T7 RNA polymerase to synthesize labeled cRNAs for microarray hybridization, the promoter sequence is added to the cDNA with another round of PCR. The promoter-tagged cDNA is purified with gel electrophoresis and amplified with one final cycle of PCR.

Key words: Single-cell microarray, cDNA amplification, Genome-wide analysis, Transcriptome, Developmental biology, Stem cell biology

1. Introduction

Regulation of gene expression is the primary mechanism by which a single set of genomic information gives rise to a complex array of diverse cell types in multicellular organisms. The completion of the genome sequencing of many organisms and the development of microarray platforms encompassing whole-genome information have provided great opportunities for analyzing genome-wide gene expression profiles in various biological settings (1). Since it is often the case that a small number of key cell types regulate important aspects of biological processes, genome-wide analysis of gene expression in a small number of cells, ideally in single cells, is a critical challenge for modern molecular biology.

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_7, © Springer Science+Business Media, LLC 2011

We have developed a polymerase chain reaction (PCR) method that faithfully amplifies global cDNA from single mammalian cells (2–4) by improving preceding methods (5–7); the improvement includes short time length of reverse transcription (RT), removal of residual primers after the first-strand cDNA synthesis, reduction of the risk for random error during PCR by performing PCR in four different tubes, use of efficient primers for PCR, reduced PCR cycle number, and use of two different primers for 5′ and 3′ end of cDNA for amplification. This method produces double-stranded cDNA with preserved sense–antisense orientation, which enables the cDNA to be easily applied to the in vitro transcription (IVT)-based biotin labeling, for commercially available high-density oligonucleotide microarrays (e.g., GeneChip series (Affymetrix)).

The success of this method was quantified by real-time PCR; in 65% of the genes examined, there was a less than twofold difference in expression levels between the amplified cDNA and original RNA, and in 89%, there was a less than fourfold difference (2). Furthermore, microarray analyses showed that the rates of false positives and negatives were as low as 3 and 6%, respectively, for genes expressed at levels of 20 copies per cell or more (45 and 30% even at five copies per cell, respectively), underscoring the method's sensitivity and accuracy (2).

This method requires no purification of RNA or cDNA from single cells and produces a large amount of double-stranded DNA in half a day. In our hands, the amplification success rate was estimated to be more than 90% for the 10 pg ES-cell total RNA and about 70% for real single cells (see Note 28) (2, 3). Because of these advantages, the method can be used to amplify many cells in parallel with reasonable time and effort; we have recently used it to simultaneously amplify cDNAs from 20 single cells per operator, and a total of more than 1,000 cDNAs in a single study (8).

The usefulness of this method has been demonstrated in studies on mouse blastocyst inner cell mass (ICM) (2), primordial germ cell (PGC) specification (8, 9), and neural progenitors in cerebral development (10). The method is, thus, applicable to a wide range of biological questions that require gene expression analysis at single-cell resolution.

2. Materials

2.1. Single Cell Isolation

1. Solution for embryo isolation: DMEM (Sigma)-BSA (1 mg/ml) (store at 4°C).
2. 0.05% Trypsin solution for cell dissociation (Invitrogen) (store at 4°C).

3. Solution for cell isolation: PBS (pH 7.2)–BSA (0.2 mg/ml) (store at 4°C).
4. Borosilicate glass capillary for mouth pipette (Sutter Instruments; O.D.: 1.0 mm; I.D.: 0.58 mm; length: 6 cm).
5. Model P-97/IVF micropipette puller for processing of glass capillary (Sutter Instruments).
6. Microforge MF-900 for processing of glass capillary (Narishige).
7. Mouth pipette is assembled as described in Subheading 3.1.
8. Stereomicroscope of magnification 50× or greater.

2.2. cDNA Synthesis and Amplification

1. Deionized, distilled water (DDW) (Invitrogen) (store at –20°C).
2. Universal primers for cDNA amplification: V1 $(dT)_{24}$ ATATGGATCCGGCGCGCCGTCGACTTTTTTTTTTTTT TTTTTTTTTTTT; V3 $(dT)_{24}$, ATATCTCGAGGGCGCGC CGGATCCTTTTTTTTTTTTTTTTTTTTTTTT; T7-V1, GGCCAGTGAATTGTAATACGACTCACTATAGG GAGGCGGATATGGATCCGGCGCGCCGTCGAC (HPLC grade) (Qiagen or Hokkaido System Science) (store at –20°C).
3. Qiagen RNase inhibitor (Qiagen, 30 U/μl) (store at –20°C).
4. Ribonuclease inhibitor (human placenta) (Takara, 40 U/μl) (store at –20°C).
5. SuperScript III reverse transcriptase (Invitrogen, 200 U/μl) (store at –20°C).
6. T4 gene 32 protein (Roche, 5–6 mg/ml) (store at –20°C).
7. Exonuclease I (Takara, 5 U/μl) (store at –20°C).
8. TdT, Recombinant (Invitrogen, 15 U/μl) (store at –20°C).
9. RNaseH (Invitrogen, 2 U/μl) (store at –20°C).
10. Takara Ex Taq Hot Start Version (Takara, 5 U/μl) (store at –20°C).
11. dATP, 100 mM (GE Healthcare) (store at –20°C).
12. Nonidet P-40 SP (Nakalai Tesque) (store at 4°C).
13. GeneAmp 10× PCR buffer II (without $MgCl_2$) (Applied Biosystems) (store at –20°C).
14. $MgCl_2$ solution (Applied Biosystems, 25 mM, included with PCR bufferII [reagent 13]) (store at –20°C).
15. DTT (Invitrogen, 0.1 M, attached with SuperScript III [reagent 5]) (store at –20°C).
16. 10× Ex Taq Buffer (Takara, attached with ExTaqHS [reagent 10]) (store at –20°C).
17. dNTP mix (Takara, 2.5 mM each of dATP, dCTP, dGTP, dTTP, included with ExTaqHS [reagent 10]) (store at –20°C).

18. 10× Exonuclease I buffer (Takara, attached with Exonuclease I [reagent 7]) (store at –20°C).
19. DNA purification kit: Qiaquick PCR Purification Kit (Qiagen).
20. Tube for first-strand cDNA synthesis: 0.5 ml thin-walled PCR tube with flat cap.
21. Tube for PCR: 8-well 0.2 ml thin-walled PCR tube band without cap.
22. Pipette tips (1,000, 200, and 10 μl).
23. Two water baths for the RT reaction.
24. PCR machine connected to a 96-well interchangeable block and 0.5-ml interchangeable block.

2.3. Template Preparation for IVT Labeling

1. Reagents for cRNA synthesis: MEGAscript T7 kit (Ambion) (store at –20°C).
2. Agarose gel for cDNA purification: UltraPure Agarose (Invitrogen).
3. Gel Extraction Kit (Qiagen).
4. Electrophoresis tank.
5. UV_{365nm} transilluminator.

2.4. Control Experiment to Establish the Protocol

1. Source of total RNA: Eukaryotic cells (~10^7 cells).
2. RNA purification kit: RNeasy Mini Kit (Qiagen).

2.5. Spike RNA

1. Template plasmids for spike RNAs: pGBS-Lys (American Type Culture Collection [ATCC], 87482), pGBS-Phe (87483), pGBS-Thr (87484), and pGBS-Dap (87486) (store at –20°C).
2. Plasmid preparation kit: Qiagen plasmid Midi/Maxi kit (Qiagen).
3. Restriction enzyme for plasmid linearization: *Not* I (store at –20°C).
4. Reagent for spike RNA synthesis: MEGAscript T3 kit (Ambion) (store at –20°C).

2.6. Real-Time PCR

1. Reagent for real-time PCR: SYBR Green PCR mastermix or Power SYBR Green PCR mastermix (Applied Biosystems) (store at 4°C).
2. Primers for genes of interest (Qiagen or Hokkaido System Science) (store at –20°C).
3. Consumable material for real-time PCR: MicroAmp Optical 384 well reaction plate (Applied Biosystems).
4. Consumable material for real-time PCR: MicroAmp Optical Adhesive Film (Applied Biosystems).
5. Equipment for real-time PCR: 7900HT Fast Real-Time PCR System (Applied Biosystems).

3. Methods

Briefly, (1) an isolated single cell is put into a tube containing cell lysis buffer without RNA purification (see Subheading 3.1); (2) the first-strand cDNA is synthesized by reverse transcriptase with the first primer V1 $(dT)_{24}$; (3) the first primer is degraded by Exonuclease I; (4) first-strand cDNA is tailed with poly (dA) by terminal deoxynucleotidyl transferase (TdT), and mRNA is degraded by RNase H; (5) poly (dA)-tailed cDNA is split into four tubes so as to reduce the risk of random error during the PCR (some genes in PCR products amplified from the same first-strand cDNA show difference in expression levels), and the second-strand cDNA is synthesized by Taq DNA polymerase with the second primer V3 $(dT)_{24}$; and (6) the first primer is again added and the cDNA is amplified with PCR, which produces double-stranded cDNA with the sense–antisense orientation preserved (see Subheading 3.2) (Fig. 1). The promoter of T7 RNA polymerase for IVT labeling is added to the cDNA by an additional 9-cycle PCR without distorting the proportion of expression levels (see Subheading 3.3) (Fig. 1).

For in-depth real-time PCR analysis of many single cells as an alternative to genome-wide analysis on microarrays, a good amount of cDNA is obtained with an additional 9-cycle PCR (see Subheading 3.4).

To establish the protocol in laboratories, we strongly recommend performing a control experiment by amplifying purified total RNA diluted to the single cell level (10 pg) and comparing the products with the cDNA synthesized from nondiluted, nonamplified RNA of the same origin (see Subheading 3.5).

To estimate the quality of cDNA amplification, a set of external RNA molecules of known copy numbers can be included (spike RNAs) (see Subheading 3.6). We use a set of spike RNAs supported by Affymetrix GeneChip microarrays: *Bacillus subtilis* mRNAs artificially tailed with poly (A) (*Lys*, *Dap*, *Phe*, and *Thr* RNAs of 1,000, 100, 20, and 5 copies per reaction, respectively). The spike RNAs can also be used as good measures for the estimation of absolute transcript copy numbers.

Finally, the gene expression levels can be measured by real-time PCR using cDNAs as templates (see Subheading 3.7).

3.1. Single Cell Isolation

The isolation and collection of single cells depends on the tissues or biopsies of interest and the equipment available (see Note 1). We here describe the protocols we used to analyze the ICM of mouse blastocysts and the PGCs or their precursors in the epiblast and extraembryonic mesoderm of gastrula embryos (embryonic day [E] 6.25–8.25).

1. Prepare micropipettes with diameters appropriate for single cell isolation. Pull a borosilicate glass capillary with a micropipette

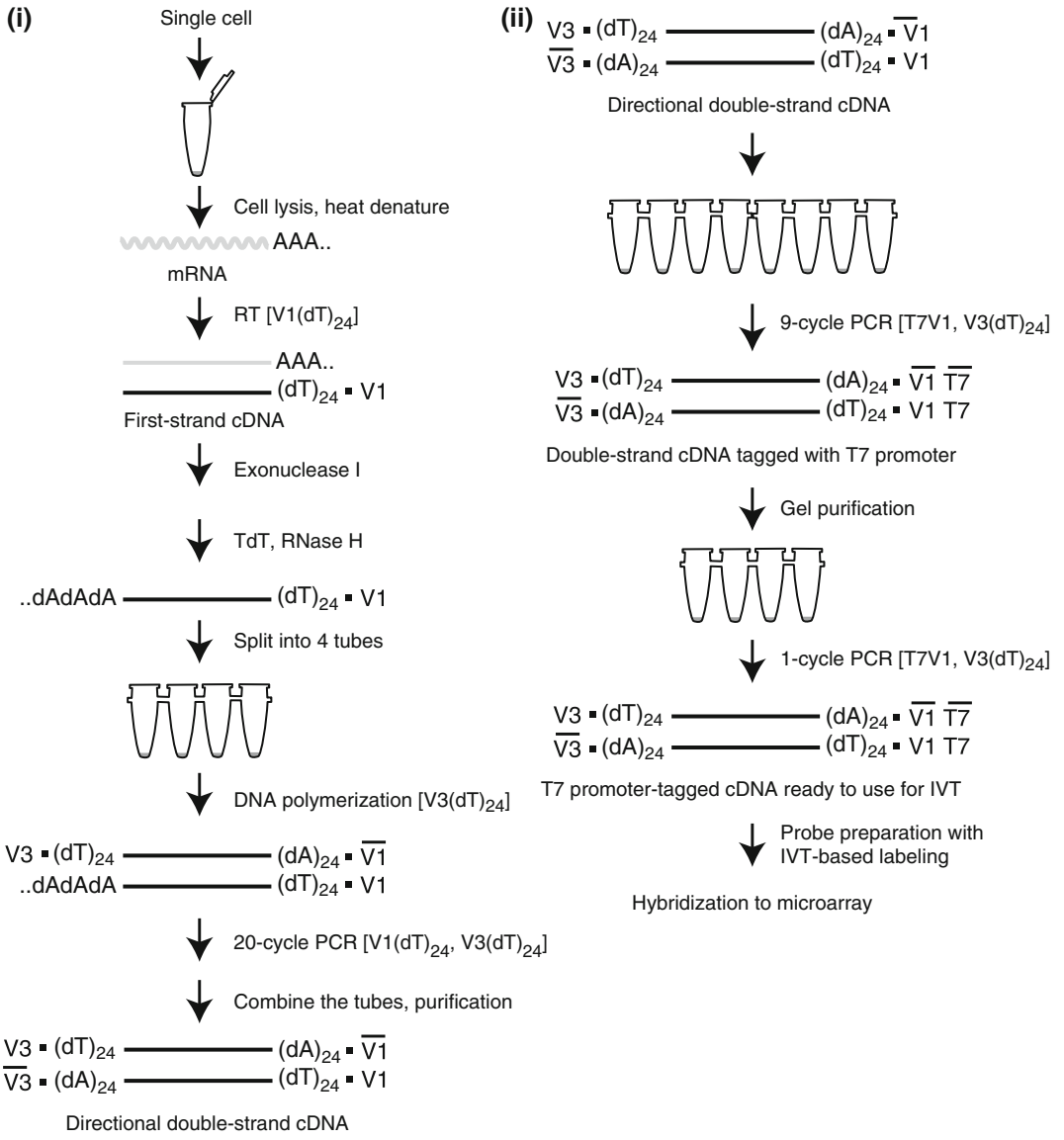


Fig. 1. Schematic representation of single-cell cDNA amplification. Shown are procedures for the initial cDNA amplification (**i**) and the subsequent template preparation for IVT labeling (**ii**).

puller; in our previous studies, we used a Sutter Instrument Model P-97/IVF with settings of P = 300, HEAT = 850, PULL = 30, VEL = 120, and TIME = 200. Cut the capillary with the microforge to make a diameter of about 30 μm, which is appropriate for the collection of single cells. Assemble the mouth pipette as described in Fig. 2.

2. Prepare the glass knives. Melt down the central portion of a borosilicate glass capillary with a gas burner, forming an uncavitated solid part. Cool the capillary at room temperature

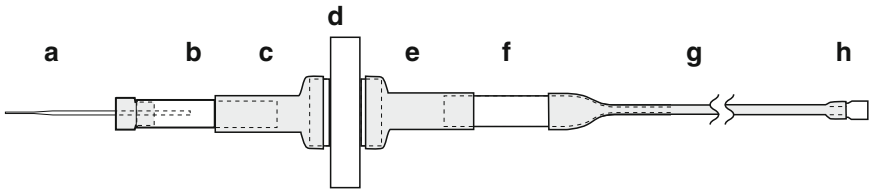


Fig. 2. Mouth pipette assembly. (**a**) Micropipette (I.D., 30–40 μm). (**b**) Adaptor for micropipette. (**c**) Silicon tube. (**d**) PVDF filter (0.45 μm). (**e**–**g**) Adapter for the filter and mouth piece (made from a Pasteur pipette and silicon tubes). (**h**) Mouth piece.

and then pull it apart with a micropipette puller as in the above settings. One capillary yields two glass knives.

3. Isolate an embryo fragment of interest by dissecting an embryo with a glass knife in DMEM–BSA.
4. Incubate the fragment in 0.05% Trypsin–EDTA at 37°C for 7 min and then gently transfer it to a drop of PBS–BSA solution (~0.3 ml). Note that the adherence junction is loosened, and avoid the scattering of cells during the transfer.
5. Dissociate the fragment into single cells by gently pipetting with the mouth pipette.
6. Collect a single cell with the mouth pipette into a sample tube containing cell lysis buffer and keep it on ice (see Notes 2–5).

3.2. cDNA Synthesis and Amplification

1. Prepare two water baths at 50 and 70°C, respectively.
2. Prepare master mix solutions without enzymes and an empty tube for RT mix as listed in Tables 1–6 and keep them on ice (see Notes 6–13).
3. Add RNase inhibitors to the cell lysis buffer and mix gently. If necessary (see Note 19), add spike RNAs. Aliquot 4.5 μl of the solution into a 0.5-ml thin-walled PCR tube (referred to as the "sample tube" hereafter) and keep it on ice for up to 2 h.
4. Prepare single cells (see Subheading 3.1).
5. Add an isolated single cell or 10 pg purified total RNA (see Subheading 3.5) into the sample tube.
6. Incubate the sample tube at 70°C for 90 s in the water bath and immediately put it on ice for 1 min (see Note 14).
7. Prepare the RT mix solution by gently mixing the enzymes (Table 2).
8. Spin down the sample tube briefly, add 0.3 μl RT mix and spin it down again.
9. Incubate the sample tube at 50°C for 5 min and at 70°C for 10 min. Then, immediately put it on ice for 1 min.
10. Add Exonuclease I to the ExoI mix solution and mix gently.
11. Spin down the sample tube briefly, add 1 μl ExoI mix, and spin the sample tube down again.

Table 1
Cell lysis buffer

Number of reactions	×1 (μl)	×20 (μl)
10× PCR buffer II (without $MgCl_2$)	0.45	9.0
$MgCl_2$ (25 mM)	0.27	5.4
NP40 (5%)	0.45	9.0
DTT (0.1 M)	0.225	4.5
Qiagen RNase inhibitor[a]	0.045	0.9
Ribonuclease inhibitor (human placenta)[a]	0.045	0.9
V1$(dT)_{24}$ primer (10 ng/μl)	0.09	1.8
dNTP mix (2.5 mM each)	0.09	1.8
Spike RNA solution[b]	×	×
DDW	2.835–×	56.7–×
Total volume	4.5	90

[a]Add RNase inhibitors just before use
[b]Add after RNase inhibitor if necessary (see Note 19)

Table 2
RT mix

Number of reactions	×1 (μl)	×20 (μl)
SuperScriptIII	0.2	4.0
Ribonuclease inhibitor (human placenta)	0.033	0.66
T4 gene 32 protein	0.067	1.34
Total volume	0.3	6.0

Chill an empty tube on ice and add enzymes just before use

Table 3
ExoI mix

Number of reactions	×1 (μl)	×20 (μl)
10× Exonuclease I buffer	0.1	2.0
DDW	0.8	16.0
Exonuclease I[a]	0.1	2.0
Total volume	1.0	20.0

[a]Add the enzyme just before use

Table 4
TdT mix

Number of reactions	×1 (μl)	×20 (μl)
10× PCR buffer II (without $MgCl_2$)	0.6	12.0
$MgCl_2$ (25 mM)	0.36	7.2
dATP (100 mM)	0.18	3.6
DDW	4.26	85.2
RNaseH[a]	0.3	6.0
TdT[a]	0.3	6.0
Total volume	6	120

[a]Add the enzymes just before use

Table 5
PCR mix I

Number of reactions	×1 (μl)	×20 (μl)
10× ExTaq buffer	7.6	152
dNTP mix (2.5 mM each)	7.6	152
V3$(dT)_{24}$ primer (1 μg/μl)	1.52	30.4
DDW	58.52	1170.4
ExTaq Hot Start Version[a]	0.76	15.2
Total	76	1520

[a]Add DNA polymerase just before use

Table 6
PCR mix II

Number of reactions	×1 (μl)	×20 (μl)
10× ExTaq buffer	7.6	152
dNTP mix (2.5 mM each)	7.6	152
V1$(dT)_{24}$ primer (1 μg/μl)	1.52	30.4
DDW	58.52	1170.4
ExTaq Hot Start Version[a]	0.76	15.2
Total	76	1520

[a]Add DNA polymerase just before use

12. Incubate the sample tube at 37°C for 30 min and at 80°C for 25 min in a thermal cycler attached to a 0.5-ml interchangeable block. Then, immediately put it on ice for 1 min.
13. Add TdT and RNase H to the TdT mix solution and mix gently.
14. Spin down the sample tube briefly, add 6 μl TdT mix, and spin the sample tube down again.
15. Incubate the sample tube at 37°C for 15 min and at 70°C for 10 min. Then, immediately put it on ice for 1 min (see Note 15).
16. Add ExTaq HS DNA polymerase to PCR mix I and II and mix gently.
17. Spin down the sample tube briefly and aliquot the content into four wells of the eight-well tube band (3 μl per well). Keep the tube band on ice.
18. Add 19 μl PCR mix I to each well and spin down the tube band briefly.
19. Perform a first round PCR consisting of a single cycle of 95°C for 3 min, 50°C for 2 min, and 72°C for 3 min. Then, immediately put the tube band on ice for 1 min.
20. Add 19 μl PCR mix II to each well, along with a drop of mineral oil, and spin down the tube band briefly.
21. Perform the second round PCR consisting of 20 cycles of 95°C for 30 s, 67°C for 1 min, and 72°C for 3 min (+6 s for each cycle), followed by a single cycle of 72°C for 10 min, and keep at 4°C.
22. Combine the PCR product (164 μl), purify it with a Qiagen PCR Purification Kit according to the manufacturer's instructions and elute it in 50 μl buffer EB. The cDNA can be stored at –20°C or –80°C. A typical gel electrophoresis pattern of the initial amplification product is shown in Fig. 3 (see Notes 16–18).
23. If necessary (see Note 20), screen the cell type of interest by the usual PCR with gene-specific primers using 0.3 μl of the cDNA as a template (primers for the real-time PCR are applicable [see Subheading 3.7]) and gel electrophoresis analysis.

3.3. Template Preparation for IVT Labeling

1. Prepare PCR mix III (Table 7) and chill on ice.
2. Aliquot 0.65 μl cDNA amplified by the 20-cycle PCR (see Subheading 3.2) into each well of the 8-well 0.2-ml tube band (total 5.2 μl) and keep it on ice.
3. Add 49.5 μl PCR mix III to each well, along with a drop of mineral oil, and spin down the tube band.
4. Perform a 9-cycle PCR consisting of 1 cycle of 95°C for 5 min 30 s, 64°C for 1 min, and 72°C for 5 min 18 s and 8 cycles of 95°C for 30 s, 67°C for 1 min, and 72°C for 5 min 18 s (+6 s for each cycle), followed by a single cycle of 72°C for 10 min (see Note 21).

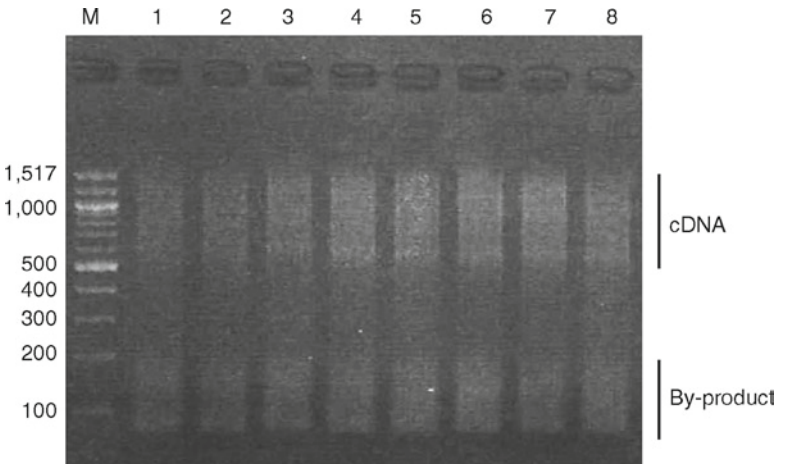


Fig. 3. Typical gel electrophoresis result of amplified cDNA. Eight samples of 10 pg ES-cell total RNA were independently amplified with 24-cycle PCR instead of 20-cycle PCR to obtain an amount of RNA sufficient to be visualized with ethidium bromide. 5 µl out of 50 µl of total cDNAs (lanes 1–8) and 100 ng of the 100-bps DNA ladder (NEB) (lane M) were loaded.

Table 7
PCR mix III

Number of reactions	×1 (µl)
10× Extaq Buffer	40
dNTP mix (2.5 mM each)	40
T7V1 primer (1 µg/µl)	8.0
$V3(dT)_{24}$ primer (1 µg/µl)	8.0
DDW	294.96
ExTaq Hot Start Version	4.0
Total volume	394.96

5. Combine the PCR product, purify it with the Qiagen PCR Purification Kit and elute it in 30 µl buffer EB.
6. Purify the cDNA with 2% agarose gel. Run electrophoresis in 0.5× TAE at 100 V for 10 min and excise the cDNA smear larger than 250 bps on a UV_{365nm} transilluminator (365 nm). Purify the DNA with a Gel Extraction Kit and elute it in 30–50 µl buffer EB (see Notes 22–24).
7. Perform the final 1-cycle PCR. Prepare PCR mix IV (Table 8) and chill on ice. Aliquot 1.9 µl gel-purified cDNA into each of four 0.2-ml PCR tubes (6 µl in total), and add 48.1 µl PCR mix IV. Perform the 1-cycle PCR consisting of 95°C for 5 min 30 s, 67°C for 1 min, and 72°C for 16 min. Combine and purify the PCR products with a Qiagen PCR Purification

Table 8
PCR mix IV

Number of reactions	×1 (μl)
10× Extaq Buffer	20
dNTP mix (2.5 mM each)	20
T7V1 primer (1 μg/μl)	4.0
V3$(dT)_{24}$ primer (1 μg/μl)	4.0
DDW	142.4
ExTaq Hot Start Version	2.0
Total volume	192.4

Kit or GeneChip Sample Cleanup Module and elute it in 30 μl buffer EB.

8. Check the template quality with nonlabel IVT. Perform IVT with a MegaScript T7 Kit at a 5 μl scale for 16 h using 1 μl of the DNA. Purify the product with an RNeasy Kit and measure the yield by UV260 absorption (1 O.D. = 40 μg/ml). A typical yield is more than 10 μg (see Notes 25, 26).
9. Check the representation of gene expression in the IVT product. Perform RT with SuperScript III according to the manufacturer's instructions using 0.5 μg IVT product as a template. Purify the RT product (IVT–RT product) with a PCR Purification Kit and elute it in 50 μl buffer EB. Compare the gene expression levels in the IVT–RT product and template cDNA by real-time PCR (see Subheading 3.7).
10. Perform microarray experiments according to the manufacturer's instructions. Typically, we apply 3–11 μl of the DNA to the Affymetrix Eukaryotic Target Preparation protocol, starting with the "Synthesis of Biotin-Labeled cRNA for Both the One-Cycle and Two-Cycle Target Labeling Assays" step. A typical yield of labeled cRNA is more than 50 μg.

3.4. Additional Amplification (Optional)

This is a scaled-down version of the procedure described in Subheading 3.3, with the PCR performed in a single tube (see Note 27).

1. Prepare PCR mix V (Table 9) and chill on ice.
2. Aliquot 0.65 μl cDNA amplified with the 20-cycle PCR (see Subheading 3.2) into a 0.2-ml PCR tube and keep on ice.
3. Add 49.5 μl PCR mix III to the sample, along with a drop of mineral oil, and spin down the PCR tube.
4. Perform a 9-cycle PCR as described in Subheading 3.3.

Table 9
PCR mix V

Number of reactions	×1 (μl)
10× Extaq Buffer	5
dNTP mix (2.5 mM each)	5
T7V1 primer (1 μg/μl)	1
V3$(dT)_{24}$ primer (1 μg/μl)	1
ExTaq Hot Start Version	0.5
DDW	36.87
Total volume	49.37

5. Purify the PCR product with a Qiagen PCR Purification Kit and elute it in 50 μl buffer EB.

3.5. Control Experiment (Optional)

1. Prepare total RNA from eukaryotic cells with an RNeasy kit according to the manufacturer's instructions. The choice of cell type will depend on the goals of research. We used an embryonic stem (ES) cell culture in (2) and obtained about 200 μg total RNA from 1×10^7 cells.
2. Dilute the total RNA with DDW serially to concentrations of 2.5 ng/μl, 250 pg/μl, and finally 25 pg/μl. Store them at −80°C. Although the diluted total RNA can be subjected to repeated freeze–thaw cycles, we advise to prepare several aliquots for each concentration.
3. Perform cDNA amplification (see Subheading 3.2) using 0.4 μl of 25 pg/μl RNA (10 pg) as the starting material. We recommend performing four to ten reactions from the same 25 pg/μl RNA pool at once.
4. As the gold standard, prepare cDNA from 1 μg of total RNA with SuperScript III reverse transcriptase according to the manufacturer's instructions. Purify the cDNA with a Qiagen PCR Purification Kit and elute it in 50 μl buffer EB.
5. Compare the gene expression levels in the amplified and nonamplified cDNA with real-time PCR as described in Subheading 3.7. Typical results are shown in Fig. 4 (see Notes 28–36).

3.6. Spike RNA (Optional)

1. Prepare plasmids encoding spike RNAs (pGIBS-Lys, -Dap, -Phe, -Thr) with a Qiagen Midi plasmid purification kit according to the manufacturer's instructions.
2. Linearize plasmids by *Not* I digestion. Perform in vitro transcription with a MEGAscript T3 kit on a 20-μl scale according

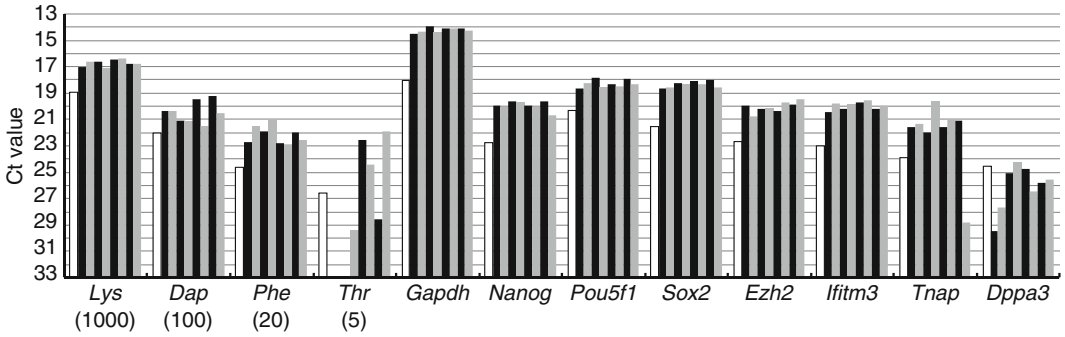


Fig. 4. Typical gene expression levels in the amplified cDNAs from 10 pg ES cell total RNA. Shown are typical results of real-time PCR analysis of cDNAs amplified from 10 pg ES cell total RNA with an initial 20-cycle PCR. The expression levels of Gapdh (13), Nanog (14, 15), Pou5f1 (16, 17), Sox2 (18), Ezh2 (19), Ifitm3 (20, 21), Tnap (22), and Dppa3 (20, 23) are shown. *Open bars* represent expression levels in 1 μg total RNA (nonamplified control), and *black* and *gray bars* represent expression levels in the independently amplified cDNA from 10 pg RNA ($N=8$). Note that the 10 pg and 1 μg RNAs are prepared from the same ES cell total RNA. The cDNA amplification and real-time PCR were performed as described in Subheadings 3.2 and 3.7, respectively.

to the manufacturer's instructions, using 0.8 μg linearized plasmids as templates. Purify the product RNA with an RNeasy kit. A typical yield is about 100 μg.

3. Prepare a master spike RNA solution from the purified RNAs so that the molar ratio of Lys:Dap:Phe:Thr equals 1,000:100:20:5. The concentration of the master solution should be as high as possible. Aliquot the master spike RNA solution and store it at −80°C. This master solution can be subjected to repeated freeze–thaw cycles (see Note 37).
4. Dilute the master solution with DDW so that the concentration of the spike RNA corresponds to 1–100 reactions per microliter (i.e., 1,125–112,500 molecules/μl). Aliquot the diluted solution for single use and store it at −80°C. This diluted solution cannot be subjected to freeze–thaw cycles.
5. Add an appropriate volume of the diluted spike RNA solution to the cell lysis buffer with RNase inhibitors. Spike RNAs can be kept in the cell lysis buffer on ice for 2 h (see Note 38).

3.7. Real-Time PCR (Optional)

1. Real-time PCR using SYBR Green is assumed here. Because the amplified cDNA is restricted to the vicinity of the 3′ end of the mRNA (~700 bps in average), primers should be designed so that the PCR product does not exceed this range. Follow general guidelines for the real-time PCR primer design; size of PCR product, 100–150; GC content, 40–60%; Tm, 60–65°C; avoidance of T, GGG, and CCC at the 3′ end (see Notes 38, 49).
2. Dilute cDNAs using Milli-Q water or DDW. The dilution rates are as follows: 1:20 for cDNAs amplified by the 20-cycle PCR (see Subheading 3.2), 1:20 for cDNAs amplified by the 9-cycle additional PCR in one tube (see Subheading 3.4),

1:10 for cDNAs synthesized from 1 μg total RNA (see Subheading 3.5), 1:40 for cDNAs synthesized using an IVT template (see Subheading 3.3), and 1:5 for the IVT–RT product (see Subheading 3.3).

3. Perform real-time PCR according to the manufacturer's instructions. We performed a 10-μl scale reaction using 2 μl of the diluted cDNA.

4. Notes

1. Any set of equipment suited to collection of single cells can be used instead.
2. The micropipette should be for single use.
3. This protocol will be applicable to early embryonic tissues that do not contain substantial extraembryonic matrices. Prolongation of the enzymatic treatment time and/or use of other enzymes such as collagenase may be required for cells embedded in extracellular matrices.
4. Take care that only one cell is collected. The single cell is not observable once it is in the reaction tube. To confirm collection of a single cell per tube, a negative control that contains only the cell isolation buffer can be included.
5. We usually carry out the experiment with two or more persons. One person collects single cells, and the others perform cDNA amplification; each operator performs amplifications for 19 single cells and one negative control, as mentioned above. It is recommended that all the cell isolation procedures be finished as quickly as possible (~2 h).
6. The usual precautions should be taken when handling RNA (e.g., keeping the clothes and bench as clean as possible, wearing rubber gloves, keeping RNase or RNase-containing materials away from the bench, and avoiding work with both RNA and RNase on the same day). Note that expiratory air is considered to contain RNase, especially during speech.
7. Water is considered to be a critical factor for an RNase-free, efficient enzymatic reaction. Water purified with an in-house system should not be used in place of purchased DDW.
8. We did not use DEPC-treated water because this could inhibit subsequent enzymatic reactions.
9. The RNase inhibitor from human placenta may be unavailable in the near future. Our preliminary experiments suggested that it can be replaced with other RNase inhibitors (e.g., the RNase inhibitor from porcine liver [Takara]).

10. It is not necessary to siliconize or autoclave the tubes and tips. Most sterilized plastic tubes and tips for single use will be suitable for this method.
11. We aliquot the nonenzymatic reagents for several times of use to reduce the risk of contamination. We store them at –20°C, except for NP-40 (4°C) and mineral oil (25°C).
12. Considering the pipetting error, we recommend preparing slightly larger volumes of the master mix solutions than required (e.g., 1.1-fold excess).
13. It is critical to keep all the master mixtures on ice until just before use.
14. The temperature of the water bath should be homogenized by agitation. A nonagitated water bath may not work for this method. We use Thermo Minder 50 mini water bath (Titec).
15. We do not store the first-strand cDNA before amplification; the nonamplified, single-stranded cDNAs from single cells run the risk of being degraded by the residual activity of Exonuclease I.
16. The average length of the cDNAs amplified with this method is approximately 700 bps, and almost all the cDNAs bear bona fide transcript ends; a BLAST search for 40 clones of amplified cDNA demonstrated that 39 clones (97.5%) had transcript ends identical to those of the nucleotide sequences registered in the NCBI database (2). Since the probes of commercially available microarrays are generally designed at the 3′ end of the transcripts, the 3′-restriction is not a major drawback for microarray and real-time PCR analysis. However, this method may not be applicable to analyses that strictly require full-length cDNAs, such as splice variant and transcription start site analyses.
17. The PCR cycle number is chosen so that the primers and dNTP are not depleted, and the cDNA amplification is stopped during the exponential phase. Therefore, the cycle number can be changed as appropriate. For 10 pg ES cellular total RNA, the cDNA is amplified exponentially with PCR of up to 28 cycles (2, 3).
18. As for the PCR machine, we use GeneAmp PCR System 9700, connected to a 96-well interchangeable block and 0.5-ml interchangeable block (Applied Biosystems).
19. We recommend including spike RNA to estimate amplification quality and absolute gene expression levels (see also Subheading 3.6 and Note 28).
20. In case that the cell type of interest is not identified with morphology and/or fluorescent reporters, such cells should be identified with PCR with gene specific primers among randomly isolated cell population.

21. The PCR cycle number is chosen so that the concentration of the product is about six times that of the template cDNA to avoid the risk of distorting the relative expression levels due to saturation of PCR. Since the amount of amplified cDNA depends on that of the initial mRNA, the number of PCR cycles may be increased if a larger amount of template cDNA is necessary to obtain a sufficient amount of IVT product.
22. Since a large gel volume will reduce the yield of cDNA, the agarose gel should be prepared as thinly as possible (e.g., 20 ml agarose solution for 110 × 60 mm gel), and the electrophoresis time should not exceed 10 min.
23. Byproducts derived from the nonspecific amplification of primers form a clear band of a size smaller than 250 bps (see Fig. 3); care should be taken not to involve the byproduct DNA in the excised gel fragment. Since the average size of the primer portion contained in the cDNA (i.e., V3-poly (dT) plus poly (dA)-V1T7) is 165 bps, removal of cDNA products less than 250 bps will result in the loss of only those gene products that are less than 85 bps long, if any (2, 3).
24. Gel purification will also be required in case researchers want to apply the cDNAs to sequencing analyses (e.g., massive parallel sequencing, EST, subtraction analyses).
25. The efficiency of nonlabeling IVT is a good indicator of the faithfulness of the amplification of the labeling IVT for the probe preparation for microarray. In our hands, if the yield of the 5-μl scale nonlabeling IVT was less than 10 μg, the proportion of gene expression levels in the IVT product was distorted. The amount of template can be increased to obtain a good amount of IVT product.
26. The final PCR is designed to remove factors derived from gel purification which potentially decrease IVT efficiency and distort expression levels in the product.
27. Purification with gel electrophoresis is not necessary as long as the cDNA is used for expression analysis with real-time PCR, which is not disturbed by the by-products.
28. The success and failure of amplification can be examined by measuring the expression level of a highly expressed housekeeping gene (e.g., *Gapdh*, about 4,000 molecules per 10 pg ES cell total RNA) or spike RNAs (see also Fig. 4). Typically, deviations of C_t values for *Gapdh* in ES cells and the Lys spike RNA (1,000 molecules per reaction) measured by real-time PCR should be less than one cycle. The average levels of *Dap* (100 copies) and *Phe* (20 copies) are proportional to the copy numbers. In our hands, the success rate estimated by such quantitative amplification of spike RNA levels was more than

90% (17/18) for 10 pg ES cell total RNA and ~75% (366/488) for real single cells (2, 3); the decline of the apparent success rate for real single cells may have resulted from possible degradation of the spike RNAs in the cell lysis buffer, which was kept on ice during the cell isolation (~1.5 h).

29. The cDNA amplification protocol comprises multiple steps, all of which must work efficiently. Therefore, in case the cDNA amplification is not as good as expected, it is critical to identify the problem step by quantitative measurements (see Notes 30–35). In our trials, insufficient amplification by PCR was often the problem point. Although we tend to suspect the enzymes in the case of an insufficient amplification, in our case, they were the least problematic factor.
30. The PCR efficiency can be estimated by picking up the samples at various cycles (e.g., 8, 12, 16, 20, 24) and measuring the housekeeping gene level; if the PCR efficiency is as expected, the expression level will be approximately doubled at each cycle. Examine whether the insufficient amplification is fully explained by the low PCR efficiency.
31. The reproducibility of PCR can be estimated by comparing the gene expression in samples amplified from identical poly-A tailed cDNA, which is prepared by combining the products of Subheading 3.2, step15. The reproducibility of any process can be similarly estimated by combining the products of the immediately preceding step.
32. Our preliminary data have suggested that this method faithfully amplifies total RNA of 10 pg to 10 ng without modification. Therefore, the control experiment can be started with a relatively large amount of total RNA.
33. The standard RT reaction with SuperScript III works well with 10 ng to 1 μg total RNA. The RT efficiency and degree of RNA degradation in 10 ng RNA may be estimated by comparing the gene expression in 10 ng and 1 μg total RNAs with real-time PCR.
34. The efficiency of Exonuclease I can be examined by analyzing the amplified cDNA with 2% agarose gel electrophoresis. If the Exonuclease I reaction is successful, two separate DNA smears will be observed: one for the cDNA (500–2,000 bp) and the other for the byproduct from the amplified residual primer V1 $(dT)_{24}$ (smaller than 200 bp) (Fig. 3). If the exonuclease reaction is less efficient, the byproduct smear will be of larger size and less efficiently discriminated from the cDNA. Because the amount of cDNA amplified by a 20-cycle PCR is too small to be visually detected in the gel by ethidium bromide, the cycle number should be increased to 24–28 cycles as shown in Fig. 3.

35. In case that the PCR efficiency is good, the efficiency of the TdT reaction can be estimated by performing the amplification method with Exonuclease I reaction omitted; the smear of the by-product derived from primers is observed even when the PCR cycle number is smaller than 20.
36. We have also developed another single-cell cDNA amplification method with a single primer producing nondirectional cDNA (11). This method comprises simpler procedures and may be an alternative means of establishing the single-cell cDNA amplification procedure for the first time.
37. The molar rate of spike RNAs is calculated based on UV_{260nm} absorption and the following parameters: the approximate sizes of spike RNAs are 1.0 kb for *Lys*, 1.84 kb for *Dap*, 1.32 kb for *Phe*, and 1.98 kb for *Thr*; 1 $O.D._{260}$ is equivalent to 40 μg/ml RNA; the molecular weight of the RNA is approximately 330/nucleotide; the Avogadro number is 6.0×10^{23}. In our previous studies (2, 8–10), we prepared master spike RNA solution containing *Lys* (4.23 μg/ml), *Dap* (7.69×10^{-1} μg/ml), *Phe* (1.1×10^{-1} μg/ml), and *Thr* (4.13×10^{-2} μg/ml), 1 μl of which corresponds to 7.69×10^{6} reactions.
38. Because the concentration of the diluted spike RNA is very low, it is critical to add the spike RNA to the well-chilled tube that contains solution with RNase inhibitors.
39. We designed the primers with the help of Primer 3 software (12) using the default parameters with the following exceptions: Product size ranges, 100–200; primer size, 18–30 mer (optimal 20 mer); primer Tm, 60–70°C (optimal 65°C); Primer GC%, 40–60%.
40. Because the size of the cDNA is relatively small (500–2,000 bps), primers should be designed so as to amplify the proximity of the 3′ end of the mRNA. Therefore, it is not always easy to design primers spanning two adjacent exons. Since only two copies of genomic DNA are contained in each single cell, the risk of detecting genomic DNA is considered negligible.

Acknowledgments

K.K. is a fellow in the Special Postdoctoral Researchers Program of RIKEN. This study was supported in part by a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan and by a PRESTO project grant from the Japan Science and Technology Agency. Tables 1–8 and Figure 3 are reproduced from *Nature Protocols*, 2007 with minor modifications (3).

References

1. Lockhart, D.J., Dong, H., Byrne, M.C., Folletie, M.T., Gallo, M.V., Chee, M.S., Mittman, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E.L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* **14**, 1675–80.
2. Kurimoto, K., Yabuta, Y., Ohinata, Y., Ono, Y., Uno, K.D., Yamada, R.G., Ueda, H.R., and Saitou, M. (2006) An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res* **34**, e42.
3. Kurimoto, K., Yabuta, Y., Ohinata, Y., and Saitou, M. (2007) Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. *Nat Protoc* **2**, 739–52.
4. Saitou, M., Yabuta, Y., and Kurimoto, K. (2008) Single-cell cDNA high-density oligonucleotide microarray analysis: detection of individual cell types and properties in complex biological processes. *Reprod Biomed Online* **16**, 26–40.
5. Brady, G., Barbara, M., and Iscove, N. (1990) Representative in vitro cDNA amplification from individual hemopoietic cells and colonies. *Methods Mol Cell Biol* **2**, 17–25.
6. Iscove, N.N., Barbara, M., Gu, M., Gibson, M., Modi, C., and Winegarden, N. (2002) Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. *Nat Biotechnol* **20**, 940–3.
7. Saito, H., Kubota, M., Roberts, R.W., Chi, Q., and Matsunami, H. (2004) RTP family members induce functional expression of mammalian odorant receptors. *Cell* **119**, 679–91.
8. Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008) Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev* **22**, 1617–35.
9. Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y., and Saitou, M. (2008) Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat Genet* **40**, 1016–22.
10. Kawaguchi, A., Ikawa, T., Kasukawa, T., Ueda, H.R., Kurimoto, K., Saitou, M., and Matsuzaki, F. (2008) Single-cell gene profiling defines differential progenitor subclasses in mammalian neurogenesis. *Development* **135**, 3113–24.
11. Yabuta, Y., Kurimoto, K., Ohinata, Y., Seki, Y., and Saitou, M. (2006) Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling. *Biol Reprod* **75**, 705–16.
12. Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365–86.
13. Seipp, S. and Buselmaier, W. (1994) Isolation of glyceraldehyde 3-phosphate dehydrogenase (Gapdh) cDNA from the distal half of mouse chromosome 16: further indication of a link between Alzheimer's disease and glycolysis. *Neurosci Lett* **182**, 91–4.
14. Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003) Functional expression cloning of nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–55.
15. Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maryama, M., Maeda, M., and Yamanaka, S. (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–42.
16. Scholer, H.R., Hatzopoulos, A.K., Balling, R., Suzuki, N., and Gruss, P. (1989) A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *EMBO J* **8**, 2543–50.
17. Okazawa, H., Okamoto, K., Ishino, F., Ishino-Kaneko, T., Takeda, S., Toyoda, Y., Muramatsu, M., and Hamada, H. (1991) The oct3 gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer. *EMBO J* **10**, 2997–3005.
18. Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17**, 126–40.
19. Laible, G., Wolf, A., Dorn, R., Reuter, G., Nislow, C., Lebersorger, A., Popkin, D., Pillus, L., and Jenuwein, T. (1997) Mammalian homologues of the Polycomb-group gene Enhancer of zeste mediate gene silencing in Drosophila heterochromatin and at *S. cerevisiae* telomeres. *EMBO J* **16**, 3219–32.
20. Saitou, M., Barton, S.C., and Surani, M.A. (2002) A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293–300.

21. Tanaka, S.S. and Matsui, Y. (2002) Developmentally regulated expression of mil-1 and mil-2, mouse interferon-induced transmembrane protein like genes, during formation and differentiation of primordial germ cells. *Mech Dev* **119**, S261–7.
22. MacGregor, G.R., Zambrowicz, B.P., and Soriano, P. (1995) Tissue non-specific alkaline phosphatase is expressed in both embryonic and extraembryonic lineages during mouse embryogenesis but is not required for migration of primordial germ cells. *Development* **121**, 1487–96.
23. Sato, M., Kimura, T., Kurokawa, K., Fujita, Y., Abe, K., Masuhara, M., Yasunaga, T., Ryo, A., Yamamoto, M., and Nakano, T. (2002) Identification of PGC7, a new gene expressed specifically in preimplantation embryos and germ cells. *Mech Dev* **113**, 91–4.

Chapter 8

Quantitation of MicroRNAs by Real-Time RT-qPCR

Caifu Chen, Ruoying Tan, Linda Wong, Richard Fekete, and Jason Halsey

Abstract

MicroRNAs (miRNAs) are ~22 nucleotide regulatory RNA molecules that play important roles in controlling developmental and physiological processes in animals and plants. Measuring the level of miRNA expression is a critical step in methods that study the regulation of biological functions and that use miRNA profiles as diagnostic markers for cancer and other diseases. Even though the quantitation of these small miRNA molecules by RT-qPCR is challenging because of their short length and sequence similarity, a number of quantitative RT-qPCR-based miRNA quantitation methods have been introduced since 2004. The most commonly used methods are stem-loop reverse transcription (RT)-based TaqMan® MicroRNA assays and arrays. The high sensitivity and specificity, large dynamic range, and simple work flow of TaqMan® MicroRNA assays and arrays have made TaqMan analysis the method of choice for miRNA expression profiling and follow-up validation. Other methods such as poly (A) tailing-based and direct RT-based SYBR miRNA assays are also discussed in this chapter.

Key words: MicroRNAs, RT-qPCR, TaqMan microRNA assays, SYBR microRNA assays, Expression profiling

1. Introduction

Small RNA genes that regulate gene expression were first described in 1993 by Ambros and his colleagues (1, 2). However, the term microRNA (miRNA) was only introduced in 2001 in a set of three articles in *Science* (3). MicroRNAs are short noncoding RNA molecules that are transcribed from the genomes of plants and animals. Their primary transcripts, pri-miRNAs, are processed into shorter stem-loop structures called pre-miRNAs and finally into ~22-nt mature miRNAs (4). To date, a total of 706 miRNA genes have been discovered in humans (http://microrna.sanger.ac.uk/sequences/) (5). MicroRNAs are highly conserved and regulate

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_8, © Springer Science+Business Media, LLC 2011

the expression of genes by primarily binding to complementary sequences in the 3′-untranslated regions (3′-UTR) of specific mRNA, although they can also bind to 5′ UTR or coding regions (1, 6–9). It is believed that miRNAs are the "micro-managers" of gene regulation and thus may have a profound impact on almost every cellular pathway.

There are a number of important reasons for detecting and quantifying miRNAs. First, miRNAs appear to control approximately one-third of human genes even though most of the actual miRNA targets remain elusive (10–12). Second, miRNAs have important biological functions. Recent studies indicate multilevel regulation of gene expression by miRNAs, which target a battery of mRNA genes (13–15), transcription factors (1, 2, 16–18), and alternative splicing regulators (19–21). Third, researchers have discovered that miRNAs are efficacious biomarkers for the classification of tumors and other cell types. In addition, they can be used to predict the outcome of many diseases because of their evolutionary conservation (4, 22, 23), unique expression signatures (24–27), relative stability (28, 29), and abundance (30). Since miRNAs are remarkably stable in body fluids (31) and show tissue- and disease-specific expression profiles (24, 32), tumor-derived miRNAs in serum or plasma can serve as novel biomarkers for the blood-based detection of human cancers (29, 31). Lastly, robust miRNA expression analysis will allow better understanding of the miRNA biogenesis, a key step in the validation of novel miRNA genes, which are discovered via sequencing analyses, and help determine their biological relevance.

There are several challenges in miRNA detection and quantitation. First, the short sequence of ~22 nucleotides makes it difficult to design conventional RT-qPCR assays or hybridization probes. Second, even though miRNAs are relatively abundant, the level of miRNA expression varies greatly, ranging from only a few to >100,000 copies per cell (27, 30). Next, miRNAs are not only short in length but also similar in sequences which make it difficult to design RT-qPCR assays or array probes. Lastly, there are three different forms of RNA molecules including pri-miRs (initial transcripts), pre-miRs (65-nt precursors) and the ~22-nt mature miRNAs in cells (4). It is important to detect and discriminate between these three types of RNA molecules.

A number of miRNA quantitation/profiling methods are currently available to researchers. These methods include miRNA microarrays (33–36), SYBR-based miRNA RT-qPCR assays (37–39), BeadArray (40), Invader Assays (41), and Padlock probe-based assays (42). The larger dynamic range of miRNA expression level makes hybridization-based methods less desirable because of inherent low sensitivity. Stem-loop RT based TaqMan® MicroRNA Assays are believed to be the gold-standard method with large dynamic range, high specificity and sensitivity (30, 43, 44).

2. Materials

2.1. RNA Preparations

1. mirVana™ miRNA Isolation Kit (Ambion).

2.1.1. Total RNA

2. 100% ethanol, ACS grade or higher quality.
3. 1× PBS chilled to 4°C, diluted from 10× PBS.

2.1.2. Cell Lysates

1. TaqMan® MicroRNA Cells-to-Ct™ Kit.
2. Less than 10^5 cultured cells, either attached to plates or in suspension.
3. 1× PBS chilled to 4°C, diluted from 10× PBS.

2.1.3. Tissues

1. RNA*later*® Solution.
2. (optional) RNA*later*®-ICE Solution.
3. *mir*Vana™ miRNA Isolation Kit.
4. 100% ethanol, ACS grade or higher quality.
5. 1X PBS chilled to 4°C, diluted from 10× PBS.
6. Liquid nitrogen, prechilled mortar and pestle and/or Polytron.

2.1.4. FFPE Samples

1. RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE.
2. 100% xylene, ACS grade or higher quality.
3. 100% ethanol, ACS grade or higher quality.
4. Microtome for tissue sectioning.
5. (If not using microtome) Equipment for cutting, and grinding or crushing, unsectioned core samples [e.g., Harris Micro-Punch™, small mortar and pestle (preferably polished agate), and liquid N_2].
6. (Optional) Centrifugal vacuum concentrator (e.g., SpeedVac), for drying deparaffinized tissue samples.

2.2. RT-qPCR Reagents

There are various methods for performing RT-qPCR, and although the underlying concept is similar, the components for each method vary slightly. Here, we outline the components needed for two methods: (1) direct RT approach followed by PCR and (2) non-direct RT approach, which requires polyadenylation of miRNAs prior to RT followed by PCR.

2.2.1. Stem-Loop RT

TaqMan® MicroRNA Assays (Applied Biosystems) is an example of the direct-RT method that utilizes a miRNA-specific stem-loop RT primer. Components for the RT reaction include:

1. 5× Stem-Loop RT Primer.
2. (Optional) Megaplex™ RT Primers for simultaneous detection of multiple miRNAs.

3. TaqMan® MicroRNA Reverse Transcription Kit includes: 50 U/μL MultiScribe™ Reverse Transcriptase, 100 mM (25 mM each) dNTP mix (with dTTP), 10× RT buffer, and 20 U/μL RNase inhibitor.
4. Purified or FirstChoice® Total RNA.
5. Nuclease-free water.
6. GeneAmp® PCR System 9700 Thermal Cycler.
7. General lab materials and equipment: nuclease-free microcentrifuge tubes, 96- or 384-well reaction plates, aerosol-resistant pipette tips, caps/seals for tubes/plates, pipettes, centrifuge, and disposable gloves.

2.2.2. Direct RT

1. 10× miRNA-specific RT Primer.
2. RT enzyme mix (Invitrogen) includes 5× buffer, 0.1 M DTT, 10 mM dNTPs, RNase OUT, and SuperScript III reverse transcriptase.
3. Purified or FirstChoice® Total RNA.
4. Nuclease-free water.
5. GeneAmp® PCR System 9700 Thermal Cycler.
6. General lab materials and equipment: nuclease-free microcentrifuge tubes, 96- or 384-well reaction plates, aerosol-resistant pipette tips, caps/seals for tubes/plates, pipettes, centrifuge, and disposable gloves.

2.2.3. Non-direct RT

NCode™ miRNA RT-qPCR System from Invitrogen is an example of nondirect RT approach, requiring poly(A)-tailing of miRNAs prior to the RT reaction. NCode™ miRNA First-Strand cDNA Synthesis Kit includes components for both polyadenylation and cDNA synthesis.

1. NCode™ miRNA First-Strand cDNA Synthesis Kit includes: 5× miRNA Reaction Buffer, 25 mM $MnCl_2$, 10 mM ATP, Poly(A) Polymerase, Annealing Buffer, SuperScript™ III RT/RNaseOut™ Enzyme Mix, 2× First-Strand Reaction Buffer, 25 μM Universal RT Primer, 10 μM Universal qPCR Primer, and DEPC-treated water.
2. Purified or FirstChoice® Total RNA.
3. 1 mM Tris–HCl, pH 8.0.
4. Heat block or water bath.
5. General lab materials and equipment.

2.2.4. Real-Time qPCR

TaqMan® MicroRNA Assays and SYBR-based qPCR assays are the most popular methods for miRNA quantitation. While the SYBR-based assays utilize a generic SYBR®-Green intercalating dye, TaqMan® qPCR requires miRNA-specific probes, which take

advantage of the fluorescent resonance energy transfer (FRET) system for detection. Here, we describe the components needed for both types of methods.

2.2.4.1. Components for TaqMan® MicroRNA Assays or Arrays Include

1. cDNA from the RT reaction.
2. (Optional) Megaplex™ PreAmp Primers for array platform.
3. (Optional) TaqMan® PreAmp Master Mix.
4. 20× TaqMan® Assay that includes forward and reverse PCR primers and TaqMan® probe.
5. TaqMan® 2× Universal PCR Master Mix, No AmpErase® UNG.
6. Nuclease-free water.
7. Applied Biosystems 7900HT Fast Real-Time PCR Instrument and User Guide.
8. ABI PRISM® 96-Well or 384-well Clear Optical Reaction Plate with barcode.
9. ABI PRISM® Optical Adhesive Covers.
10. TaqMan® Human or Rodent MicroRNA Array Set v2.

2.2.4.2. Components for Invitrogen's SYBR®-Based NCode™ miRNA qPCR Include

1. cDNA from the RT reaction.
2. 10 μM Universal qPCR Primer.
3. SYBR® GreenER™ qPCR SuperMix Universal that includes 2× SYBR® GreenER™ qPCR SuperMix and ROX reference dye.
4. 10 μM Custom Forward qPCR Primer: the forward primer is specific for the miRNA sequence and is ordered separately by the user. Ordering DNA oligo that is identical to the entire mature miRNA sequence is recommended.
5. qPCR instrument and appropriate PCR tubes/plates for instrument.

3. Methods

3.1. RNA Preparations

3.1.1. Total RNA

1. A general protocol for total RNA isolation (large and small RNA) on a single filter is given here. This protocol uses higher ethanol concentrations to bind all RNA species at one time to the filter. Sample-specific changes will be provided below in the appropriate section.
2. Using mirVana™ miRNA Isolation Kit, add 1/10 volume of the miRNA Homogenate Additive to the lysate (homogenate) and mix well, and place on ice for 10 min.

3. Add a volume of acid-phenol:chloroform that is equal to the volume of the original lysate (homogenate) and vortex for 1 min.
4. Centrifuge for 5 min at maximum speed (10,000 × *g*), then remove the aqueous phase (upper phase) without disturbing the organic phase (lower phase).
5. Add 1.25 volumes of 100% ethanol to the aqueous phase, mix and pass this through the filter cartridge by centrifuging for approximately 15 s at 10,000 × *g*. Discard the flow-through.
6. Pass 700 µL of miRNA Wash Solution 1 through the filter cartridge by centrifuging for 10 s and discard the flow-through.
7. Pass 500 µL of Wash Solution 2/3 through the filter and discard the flow-through. Repeat this step.
8. Spin the filter for 1 min to remove residual wash solutions.
9. Place the filter into a new collection tube and add 100 µL of either elution solution or nuclease-free water which has been heated to 95°C, to the center of the filter. Spin for 30 s at 10,000 × *g* and collect the eluate, which contains total RNA.

3.1.2. Cell Lysate

1. Either suspension or adherent cells may be lysed and directly analyzed by RT-qPCR using the following method which uses TaqMan® MicroRNA Cells-to-Ct™ Kit. Cells may be lysed in tubes (after trypsinization) or lysed directly in culture plates. Keep cells on ice when not manipulating.
2. Remove media from cells attached in culture plates and wash cells with approximately 100 µL of cold 1× PBS. Remove PBS without disturbing cells.
3. To detach adherent cells, remove the culture medium and rinse once with PBS. Trypsinize the cells to detach them and then inactivate the trypsin. Pellet between 10 and 10^5 cells at a centrifugal force appropriate for the specific cell line and discard the culture medium (or Trypsin solution). Wash the cells by resuspending them in 1 mL of 1× PBS and repellet the cells.
4. Dilute DNase into Lysis Solution at a ratio of 1:100. Add 50 µL of Lysis Solution mixture to the cells (in a maximum of 5 µL of 1× PBS) and pipette up and down five times. Set pipette to 35 µL during mixing to ensure that no bubbles are formed.
5. Incubate the reaction mixture for 8 min at room temperature.
6. Add 5 µL of Stop Solution (ensure that all Stop Solution enters lysis solution by touching the tip to the bottom of the lysed cell mixture), then pipette up and down five times. Incubate at room temperature for 2 min (Do not allow lysate to remain longer than 20 min after addition of Stop Solution). Lysates are stable on ice for up to 2 h and for up to 5 months at –20 or –80°C.

7. Perform RT reactions (below) and substitute lysate for pure RNA without the need for RNA quantitation.

3.1.3. Tissues

1. The following protocol uses the mirVana™ miRNA Isolation Kit to obtain high-quality small RNA. It is very important to minimize the time between tissue collection and RNase inactivation (step 4 below).
2. Collect the tissue and remove as much extraneous material as possible. Red blood cells can be removed by washing the tissue with cold 1× PBS.
3. Cut the tissue into small enough pieces for physical disruption and obtain a sample between 0.5 and 250 mg.
4. (Optional) Inactivate the RNases by either of the following:
   (a) Placing the samples into RNAlater according to the protocol. To allow proper perfusion, tissue must be cut into pieces with at least one dimension smaller than 0.5 cm. Remove tissue from RNAlater and blot to remove excess before proceeding to subsequent steps.
   (b) Freeze sample in liquid nitrogen. Samples must be small enough to freeze within a few seconds, store them at −70°C or colder temperatures.
5. Place tissue sample into approximately 10 volumes per tissue mass of Lysis/Binding Buffer in a vessel appropriate for homogenization.
6. While keeping the samples cold, homogenize the tissue using a ground-glass homogenizer, plastic pestle (for small pieces), bead beater, or rotor-stator homogenizer. Ensure that no visible clumps of tissue remain.
7. Proceed with Total RNA (Subheading 3.1.1) isolation procedures given above.

3.1.4. FFPE Samples

1. Total RNA (small and large) is recovered by using the RecoverAll™ Isolation Kit. RNA fragmentation from fixation/embedding seems to have an endpoint of approximately 80 nucleotides generally leaving small RNA, like miRNA, intact. MiRNA recovery can be maximized by ensuring FFPE slices (greater than 10 μm or one cell thick) and by limiting the time during deparaffinization in xylene and ethanol washes.
2. Collect approximately 80 μm of section (e.g., 4×20 μm sections) or less than 35 mg of unsectioned core. Grind in liquid nitrogen and place in 1 mL 100% xylene. Vortex and heat to 50°C for 3 min.
3. Centrifuge for 2 min at maximum speed and remove the xylene. The pellet is usually clear and difficult to see. Some xylene can be left to ensure that the pellet is not lost.

4. Add 1 mL of 100% ethanol to the pellet (which should turn opaque) and vortex. Centrifuge for 2 min at maximum speed. Repeat ethanol wash and ensure that the pellet is not disturbed. Recentrifuge and remove as much ethanol as possible.
5. Remove residual ethanol by air drying for 15–45 min or by vacuum centrifugation for larger sections with medium heat for 20 min or with low heat for 20–40 min.
6. Add 100–200 µL of digestion buffer (depending on sample size) with 4 µL of Protease and incubate at 50°C for 15 min and then at 80°C for 15 min.
7. For every 100 µL of digestion buffer added above, add 120 µL of Isolation Additive and 275 µL of 100% ethanol. Then, mix by pipetting up and down.
8. Place the samples into a filter cartridge. To avoid clogging the filter, do not transfer undigested pieces of tissue. Centrifuge at 10,000 × *g* for 30 s.
9. Discard flow-through and wash the filter with 700 µL of Wash 1 by centrifuging at 10,000 × *g* for 30 s.
10. Discard flow-through and wash the filter with 500 µL of Wash 2/3 by centrifuging at 10,000 × *g* for 30 s.
11. Discard flow-through and centrifuge at 10,000 × *g* for 30 s to remove remaining wash solutions.
12. Add 60 µL of DNase mix (6 µL of 10× DNase buffer, 4 µL of DNase, 50 µL of nuclease-free water) to the center of the filter cartridge, cap, and incubate at room temperature for 30 min.
13. Add 700 µL of Wash 1 to the filter cartridge and incubate at room temperature for 1 min. Centrifuge at 10,000 × *g* for 30 s.
14. Discard flow-through and wash the filter with 500 µL of Wash 2/3 by centrifuging at 10,000 × *g* for 30 s. Repeat this step.
15. Discard flow-through and centrifuge at 10,000 × *g* for 30 s to remove remaining wash solutions.
16. Transfer to a new collection tube and apply 60 µL of room temperature Elution Solution or nuclease-free water to the center of the filter. Incubate at room temperature for 1 min, then centrifuge at maximum speed for 1 min to recover the miRNA in the eluate.

3.2. Individual TaqMan® MicroRNA Assays (See Note 1)

3.2.1. Stem-Loop RT

1. Thaw components of TaqMan® MicroRNA Reverse Transcription Kit on ice.
2. In a polypropylene tube, prepare RT enzyme mix by multiplying the following amounts by the number of RT reactions needed: 7 µL RT enzyme mix contains 4.16 µL of nuclease-free

water, 0.15 μL of 100 (25 each) mM dNTP (with dTTP), 1.00 μL of MultiScribe Reverse Transcriptase (50 U/μL), 1.50 μL of 10× RT buffer, and 0.19 μL of RNase inhibitor (20 U/μL).

3. Briefly centrifuge RT enzyme mix.
4. Dilute RNA samples to proper concentration 1–10 ng/μL with 0.1× TE (pH 8.0).
5. Aliquot 7 μL of RT enzyme mix per well in a 96- or 384-well plate.
6. Add 3 μL of 5× miRNA-specific RT primer per well in a 96- or 384-well plate.
7. Add 5 μL RNA sample per well in a 96- or 384-well plate.
8. Seal the plate and mix gently. Centrifuge briefly.
9. Incubate the plate on ice for 5 min.
10. Set up RT run condition in a Thermocyler. Incubate at 16°C for 30 min, 42°C for 30 min, 85°C for 5 min, and 4°C on hold.
11. Load the plate into the thermal cycler and start the RT reaction.
12. Store the reverse transcription at –20°C if it is not used immediately.

3.2.2. Real-Time qPCR

Three to four replicates of PCRs per miRNA genes are recommended.

1. For each miRNA assayed, aliquot 10 μL of 2× Universal TaqMan® Master Mix, 1 μL of 20× TaqMan® MicroRNA Assay mix, 1.3 μL of RT product, and 7.7 μL of nuclease-free water into each of three or four wells in a 96- or 384-well plate.
2. Seal the plate with an optical adhesive cover. Centrifuge the plate briefly.
3. Refer to the appropriate instrument user guide on how to configure the plate document.
   (a) Applied Biosystems 7900HT Fast Real-Time PCR System and SDS Enterprise Database User Guide.
   (b) Applied Biosystems 7300/7500/7500 Fast Real-Time PCR System and SDS Enterprise Database User Guide.
4. Create plate document with following parameters:
   (a) Run Mode : 9600 emulation (default).
   (b) Sample Volume: 20 μL.
   (c) Thermal cycling: 95°C for 10 min, 40 cycles of (95°C for 15 s and 60°C for 60 s).

5. Open the plate document that corresponds to the reaction plate.
6. Load the reaction plate into the instrument and start the run.

3.3. TaqMan® MicroRNA Arrays

See Note 2 for schematic description of TaqMan® MicroRNA Arrays work flow and Note 3 for an example.

3.3.1. Megaplex RT

1. Thaw the following on ice: Megaplex RT Primers, TaqMan® MicroRNA Reverse Transcription Kit components, and $MgCl_2$ (supplied with the Megaplex RT Primers).
2. In a polypropylene tube, prepare the Megaplex RT enzyme mix by combining the following reagents and scale their volume to the desired number of RT reactions. If multiple RT reactions are to be run, a stock solution may be prepared by multiplying each volume in the table by the number of RT reactions anticipated.

Components	Magaplex RT reaction (μL)
Megaplex RT primers (10×)	0.75
dNTP with dTTP (100 mM total)	0.2
MultiScribe Reverse Transcriptase (50 U/μL)	1.5
10× RT Buffer	0.75
MgCl2 (25 mM)	0.9
RNase inhibitor (20 U/μL)	0.1
Nuclease-free water	0.3
Total	4.5

3. Mix gently. Centrifuge briefly to bring solution to the bottom of the tube.
4. Aliquot 4.5 μL of Megaplex RT enzyme mix per well in a 96- or 384-well plate.
5. Add 3 μL total RNA (1–350 ng) to each well of a 96- or 384-well plate.
6. Seal the plate and mix gently. Centrifuge to bring the solutions to the bottom of the wells.
7. Incubate the plate on ice for 5 min.
8. Set up the run methods using the following conditions:
   (a) Ramp speed or mode: use 9700 standard ramp speed for the Applied Biosystems 9700 Thermocyler.
   (b) Reaction volume (μL): 7.5.
9. Set up the RT incubation and inactivation protocol on a Thermocyler. Incubate for 40 cycles of 16°C for 2 min, 42°C for 1 min, and 50°C for 1 s. Then, inactivate RT at 85°C for 5 min. The hold temperature should be 4°C.

10. Load the plate into the thermal cycler. Start the RT run.
11. After the RT reaction, store the reverse transcription at –20°C if it is not used immediately.

3.3.2. Megaplex Preamplification (Optional, If RNA Is Limited)

1. Thaw the following on ice: Megaplex Preamplification Primers and 2× TaqMan® Preamp Master Mix.
2. Add 12.5 µL of 2× TaqMan® Preamp Master Mix to each well of the 96- or 384-well plate.
3. Transfer 2.5 µL of 10× Megaplex Preamplification Primers to each well of the 96- or 384-well plate.
4. Transfer 2.5 µL of RT product to each well of the 96- or 384-well plate.
5. Add 7.5 µL nuclease-free water to each well of the 96- or 384-well plate.
6. Seal the plate, mix gently and then centrifuge briefly.
7. Incubate the plate on ice for 5 min.
8. Set up the run methods using the following conditions:
   (a) Ramp speed or mode: use the Standard Ramp Speed on Applied Biosystems' 9700 Thermocyler.
   (b) Reaction volume: 25 µL.
   (c) Incubate at 95°C for 10 min, 55°C for 2 min, and 72°C for 2 min. Then, incubate for 12 cycles of 95°C for 15 s and 60°C for 4 min. Inactivate the polymerase at 99.9°C for 10 min. Finally, place at 4°C on hold.
9. Load the plate into the thermal cycler. Start the PreAmp PCR run.
10. After amplification, add 75 µL 0.1× TE (pH 8.0) to each 25 µL of preamplification products. The diluted preamplification products can be used directly to run real-time PCR, or stored at –20°C for at least 1 month.

3.3.3. Real-Time qPCR

1. Prepare the qPCR reaction mix in a polypropylene tube.
   (a) Add 450 µL TaqMan® Universal PCR Master Mix.
   (b) Add 444 (Megaplex RT only) or 441 (Megaplex preamplification) µL of nuclease-free water.
   (c) Add 6 µL Megaplex RT products or 9 µL diluted preamplification products.
2. Load 100 µL of the PCR reaction mix into each port of the TaqMan® MicroRNA Array. For detailed information on how to load, seal, and run the array, please refer to the Applied Biosystems TaqMan® Array User Bulletin.
3. Spin and seal the array.
4. Create plate document with following parameters.

(a) Thermal cycling conditions: 50°C for 2 min, 94.5°C for 10 min followed by 40 cycles of 97°C for 30 s and 59.7°C for 60 s.

5. In the SDS software, open the plate document that corresponds to the reaction plate.
6. Load the reaction plate into the instrument. Start the run. For detailed information on how to load, seal, and run the array, please refer to the Applied Biosystems TaqMan® Array User Bulletin.

3.4. SYBR-Based MicroRNA Assays

SYBR-based chemistry is an alternative to the TaqMan® miRNA assay chemistry (described above) for real-time detection of miRNAs. Similar to the TaqMan® miRNA detection method, SYBR-based miRNA assays involve a two-step reaction process: reverse transcription (RT) and PCR. There are two general methods for RT: the first is direct RT, in which cDNA is reverse-transcribed directly from the miRNAs, and the second involves polyadenylation of the miRNAs prior to reverse transcription. The unique challenge in the design of miRNA assays, regardless of whether it is TaqMan® or SYBR-based, is attributed to the short nature of the miRNA sequence length. A common feature in all methods, therefore, is the addition of a universal primer sequence that is incorporated into the RT primer to extend the length of the cDNA molecule during reverse transcription. The cDNA is then amplified with a miRNA-specific primer on one end and a universal primer on the other end in real-time PCR. Some methods introduce a modification, such as locked nucleic acid (LNA) base substitution, to the miRNA-specific PCR primer to enhance the quality of the assays (37).

In contrast to TaqMan® chemistry, which uses a specific probe and the fluorescent resonance energy transfer (FRET) system, SYBR-based assays rely on the enhanced fluorescence of SYBR Green dye when it intercalates into the double-stranded (ds) DNA PCR product. Because SYBR Green dye binds to all dsDNA, the biggest disadvantage to SYBR-based assays is the production of false-positive signals from nonspecific reaction products caused by primer dimers or DNA contamination.

3.4.1. Direct RT

Direct RT for SYBR-based assays is similar to the stem-loop RT of the TaqMan® MicroRNA Assays. MiRNAs in the total RNA sample are converted to cDNAs with the RT enzyme and miRNA-specific RT primers. Unlike the TaqMan® MicroRNA Assays where a novel stem-loop RT primer is used, the RT primers for SYBR-based assays are linear (37, 39) and therefore may not offer the ability to discriminate between the biologically relevant mature miRNA and the longer precursor form from which the mature miRNA is derived. Refer to Subheading 3.2 for method on TaqMan® MicroRNA Assays, which uses the direct RT approach.

3.4.2. Nondirect RT

There are a number of SYBR-based miRNA assays that utilize a more generic, nondirect RT approach (Invitrogen's NCode™ miRNA RT-qPCR System, http://www.invitrogen.com) (38). These methods require poly (A)-tailing of the miRNAs in the sample as the first step using poly (A) polymerase since miRNAs are not polyadenylated in nature. Reverse transcription is then performed with an oligo-dT-adaptor RT primer and RT enzyme to transcribe the cDNA. Incorporation of the oligo-dT-adaptor to the miRNAs during reverse transcription extends the length of the miRNAs, allowing the design of a universal reverse primer for real-time PCR. The forward primer is miRNA-specific and is generally designed to the entire length of the miRNA. Fluorescent signal for detection is through binding of the SYBR®-Green dye to the amplified PCR product. Companies that offer the nondirect RT method have simplified the work flow so that polyadenylation and RT may be achieved in the same reaction tube with optimized kits that include both the poly(A) polymerase and reverse transcription enzyme.

We outline a method for nondirect RT requiring poly(A) tailing of miRNAs, followed by SYBR-based qPCR for the quantitation of miRNAs. Invitrogen's NCode™ miRNA RT-qPCR System is used as an example.

3.4.2.1. Poly(A) Tailing

The purpose of this procedure is to add poly(A) tails to the miRNAs in the total RNA sample.

1. Use 10 ng to 2.5 μg of FirstChoice® Total RNA depending on the abundance of the miRNAs. The optimal sample range is 100 ng to 1 μg of total RNA. Note that isolating small RNA molecules from total RNA is not required. Enrichment of small RNAs may be considered for extremely low-abundant miRNAs; however, this process may limit the detection of some miRNAs depending on the small RNA size cutoff in the enrichment process. Refer to PureLink™ miRNA Isolation Kit and flashPAGE™ Fractionator System if enrichment of small RNAs is desired.
2. Dilute the 10 mM ATP in 1 mM Tris–HCl, pH 8.0, according to the manufacturer's manual.
3. Combine the following components in a nuclease-free microcentrifuge tube at room temperature. For multiple reactions, prepare a master mix of common components.

Components	Volume (μL)
RNA	10-2,500 (ng)
5× miRNA Reaction Buffer	5.0
25 mM $MnCl_2$	2.5
Diluted ATP (from step 2)	1.0
Poly(A) Polymerase	0.5
DEPC-treated water	to 25.0

4. Mix gently and centrifuge to bring solution to the bottom of the tube.
5. Incubate the tube in a heat block or water bath at 37°C for 15 min.
6. Proceed immediately to first-strand cDNA Synthesis.

3.4.2.2. RT

The purpose of this step is to reverse-transcribe the polyadenylated miRNA to generate first-strand cDNA.

1. Add the following components into a 0.2-mL nuclease-free reaction tube or 96-well reaction plate:

Components	Volume (μL)
Polyadenylated RNA (from step 5)	4
Annealing buffer	1
Universal RT Primer (25 μM)	3
Total volume	8

2. Incubate the tube at 65°C for 5 min.
3. Place the tube on ice for 1 min.
4. Add the following components to the tube for a final reaction volume of 20 μL:

Components	Volume (μL)
2× First-strand reaction buffer	10
SuperScript™ III RT/RNaseOut™ Enzyme Mix	2

5. Mix gently and centrifuge to bring solution to the bottom of the tube or plate.
6. Transfer the tube to a thermal cycler and incubate at 50°C for 50 min and then 85°C for 5 min to stop the reaction.
7. Chill the reaction on ice.
8. Store aliquots at –20°C or proceed directly to qPCR.

3.4.2.3. qPCR

The purpose of this step is to amplify the miRNA of interest using qPCR primers and DNA polymerase. Detection is possible through the binding of the SYBR® Green dye to the amplified product.

1. Dilute the cDNA (from RT step 7) 1:10 in nuclease-free water. Use 5 μL per 50 μL reaction.
2. Add the following components to each nuclease-free PCR plate well (reaction volume may be scaled down to 20 μL for 384-well plates).

Components	Volume (μL)
SYBR® GreenER™ qPCR SuperMix	25
Forward Primer, 10 μM	1
Universal qPCR Primer, 10 μM	1
ROX Reference Dye	1
Template (diluted 1:10 from qPCR step 1)	5
DEPC-treated water	to 50

3. Seal the reaction plate.
4. Mix gently and centrifuge to bring solution to the bottom of the tube or plate.
5. Load the reaction plate into Applied Biosystems 7900HT Fast Real-Time PCR instrument.
6. Run PCR using the following condition:

   50°C for 2 min; 95°C for 10 min; then 40 cycles of (95°C for 15 s and 57°C for 60 s).
7. Refer to Subheading 4 for RT-qPCR data analysis.

3.5. Data Analysis

The RT-qPCR can be used for either absolute or relative quantitation. The absolute quantity of a molecule can be determined by creating a standard curve with a molecule of known concentration and comparing this to unknown samples of interest (http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_042176.pdf). However, due to variations in starting sample quantity, sample quality and variable PCR efficiency, one of the most widely used methods for expression analysis is relative quantification. This method allows comparison of target molecules, which may vary because of different treatments or different tissue type, to a reference target(s) which does not vary. This reference may be the volume or mass of sample (such as mg of tissue or RNA); however, it is more effective to use an endogenous control (covered below). Relative quantification is also most effective when these controls or reference molecules are subjected to identical conditions (purification method, storage, etc.) as target molecules are physically similar (e.g., size) and are detected using a similar method (e.g., RT-qPCR with stem-loop primers). Using a control that undergoes the same detection process will allow for factors that may affect that detection method (such as inhibitors).

3.5.1. Selection of Endogenous Controls

One of the most accurate methods for relative expression analysis is the use of endogenous controls. An ideal endogenous control generally demonstrates gene expression that is relatively constant and highly abundant across tissues and cell types. However, one must still validate the chosen endogenous control. Ideally, a set of controls is chosen for the target cell, tissue, or treatment, as no

single control can serve as a universal endogenous control for all experimental conditions.

1. Identify a set of genes (e.g., miRNA, snoRNA, snRNA) from previous data, which demonstrate the least variation in the tissues, samples, cell lines, and experimental conditions (optional) in your study. This may be from previously published information regarding your samples or regarding candidate endogenous controls using other samples (http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_044972.pdf) (45, 46).
2. Screen these molecules using a serial dilution of your prepared samples (e.g., purified miRNA) and select candidates that provide the best assay linearity (e.g., $R^2 > 0.95$) and highest abundance.
3. Select candidates that demonstrate the best no-template control signals (NTC; e.g., Ct > 38).
4. Screen molecules using as many different samples and (optionally) treatments as will be tested.
5. Select candidates that provide least variation across samples.
6. (Optional) Compare candidates using known controls of other types such as mRNA, rRNA, or protein level and select candidates that demonstrate the best correlation.
7. (Optional) Ideally more than one endogenous control will be used to ensure the least variation due to untested samples and conditions.

3.5.2. Data Normalization

Many factors in quantitative PCR, such as the quantity and quality of the RNA and RT efficiency, contribute to the variation in gene expression. Expression data, however, can be normalized using endogenous control genes to correct for this variation. To achieve effective data normalization:

1. Identify the appropriate endogenous controls. This will often require screening for endogenous controls that perform best in the specific samples under investigation since a universally valid control gene does not exist (45, 46). A number of papers have reported the benefit of using multiple control genes for data normalization, especially when reporting small differences in expression levels (47).
2. After several candidate endogenous controls have been identified, normalize the Ct values using the averaged Ct of the endogenous controls. Differential expression may be reported as:
   (a) ΔCt (miRNA Ct – averaged endogenous control Ct)
   (b) ΔΔCt ($\Delta Ct_{(sample1)} - \Delta Ct_{(sample\ 2)}$) relative to a reference sample
   (c) Fold change ($2^{\Delta\Delta Ct}$).

TaqMan® MicroRNA Assays were designed for human, mouse, and rat small-nuclear and small-nucleolar RNAs as endogenous controls to normalize expression data for TaqMan® MicroRNA Assays. These control assays have been evaluated across a large number of normal tissues and the NCI-60 cell lines. Endogenous control assays and supporting literature are available at http://mirna.appliedbiosystems.com (also see Note 4).

3.5.3. Calculation of Fold Changes

One of the simplest and most common methods to calculate fold changes requires the use of an endogenous control (see above) and a reference sample. The reference sample may be untreated samples or any sample to which a comparison is desired. For example, "there is fivefold more let-7a in pancreas than in heart." This method is also referred to as the ΔΔCt method. For fold-change analysis, the replicate number of samples is an important factor to consider because standard deviation (SD) is required to tell if the observed fold change is statistically significant. As the number of replicates (samples and qPCR) increases, variability in the standard deviation decreases and approaches the actual variance of the system. In general, as fold changes decrease a greater number of replicates are recommended.

1. Obtain expression data for miRNAs and endogenous controls from experimental samples (e.g., treated cells) as well as from reference samples (e.g., untreated cells).
2. Calculate ΔCt for the experimental and reference samples using the following formula: $\Delta Ct = Ct_{miRNA} - Ct_{endogenous\ control}$.
3. Calculate standard deviations (SD) for each ΔCt (experimental and reference samples).
4. Calculate ΔΔCt by using the following formula: $\Delta\Delta Ct = \Delta Ct_{experimental\ sample} - \Delta Ct_{reference\ sample}$.
5. The SD for the ΔΔCt will be equal to the SD from the ΔCt of the experimental sample since the reference sample is treated as an arbitrary constant.
6. Calculate fold change by using the following formula: fold change $= 2^{-\Delta\Delta Ct}$.
7. Incorporate SD into fold change by expressing the fold change as a range. This is done by incorporating the SD into the above formula: $2^{-\Delta\Delta Ct + SD}$ and $2^{-\Delta\Delta Ct - SD}$.

3.5.4. Clustering Analysis

Expression of miRNAs is often characterized based on comparison between wild-type (or untreated) vs. mutant (or treated) samples. Biological duplicates of each sample type are highly recommended, and the expression data are subject to a series of adjustments and filtering before hierarchical clustering. Assays with Ct values greater than 35 (32 for TLDAs) are treated as 35 (32 for TLDAs) because these miRNAs are nearly undetectable.

The most stable set of endogenous control genes such as U6, snoRNAs, or miRNAs are selected as internal references based on geNorm (47), and ΔCt between the Ct of each miRNA and the average of these references for each sample is calculated. Both miRNAs and samples are clustered using the average linkage method under the correlation similarity metric. Examples of clustering analysis software include Dr. Eisen's "Cluster" (48) and Real-Time StatMiner™ from Integromics Inc.

4. Notes

1. Schematic of TaqMan® MicroRNA assays: TaqMan®-based real-time quantification of miRNA includes two steps: stem-loop RT and real-time PCR. Stem-loop RT primers bind to the 3′ portion of miRNA molecules and are reverse-transcribed with reverse transcriptase. Then, the RT product is quantified using conventional TaqMan® PCR that includes miRNA-specific forward primer, reverse primer, and a dye-labeled TaqMan® probe (30).
2. Schematic of TaqMan® MicroRNA Arrays: In MegaPlex RT step 2, miRNAs are converted into single-stranded cDNAs using the TaqMan® MicroRNA Reverse Transcription Kit and Megaplex RT Primers. In preamplification step 3, MegaPlex RT products are amplified to increase the quantity of desired cDNAs for gene expression analysis in real-time PCR. In the real-time PCR step 4–5, the DNA polymerase amplifies the target cDNA using sequence-specific primers and probe on the TaqMan® Array (step 4). The presence of the miRNA molecules is detected in real time through cleavage of the TaqMan® probe by 5′ nuclease activity (step 5).
3. An example – TaqMan® MicroRNA Arrays. Real-time quantitative PCR is a popular and powerful method used for miRNA expression profiling. Applied Biosystems has developed Megaplex Pools with matching TaqMan® MicroRNA Arrays that can simultaneously detect up to 380 miRNAs per pool (refer to Subheading 3.1.2). TaqMan® MicroRNA Assays designed for small-nuclear or small-nucleolar RNAs that have previously been evaluated for abundant and stable expression were incorporated into the TaqMan® MicroRNA Arrays as endogenous controls for data normalization. Real-Time StatMiner Software from Integromics (http://www.integromics.com) is the preferred and recommended data analysis tool for TaqMan® MicroRNA Array data analysis for multiple reasons. Raw Ct values from Applied Biosystems real-time instruments can be imported directly into the software. The software provides

relative quantification results using advanced statistics powered by R (http://www.r-project.org) and Bioconductor (http://www.bioconductor.org) algorithms. Real-Time StatMiner uses two common mathematical algorithms, geNorm (47) and NormFinder (49), for determining the most stable endogenous controls for data normalization. Real-Time StatMiner computes fold change between samples or conditions identifying genes that show significant differences and allows interactive visualization at the same time. In addition, rather than using the more common ΔΔCt method, which only allows direct sample-to-sample comparison, Real-Time StatMiner incorporates the empirical Bayes method to allow grouping of samples (i.e., according to treatment or dose). Lastly, Real-Time StatMiner provides an intuitive user interface with a step-by-step guide through the analysis process and the flexibility to navigate within the work flow. The following is the recommended procedure for data analysis of TaqMan® MicroRNA Arrays.

(a) Transfer the SDS files from the real-time instrument to an RQ study. Analyze the SDS data using Automatic Baseline and Manual Ct set to 0.2.

(b) Evaluate the amplification plots. Adjust the baseline and threshold settings if necessary. It is important, however, to apply the same threshold setting across all samples or arrays within a study.

(c) Review the gene expression plot. View the amplification plots in the Plate, Detector, or Sample view for SDS 2.3.

(d) In the well table or results table, review Ct values for each well for replicate reactions and omit justified outliers.

(e) Export raw Ct values into Real-Time StatMiner Software (Raw Ct values are exported from the Plate Centric View for SDS 2.3) and follow the steps guided by Real-Time StatMiner Software for the rest of the procedure below.

(f) Perform quality control of biological replicates.

(g) Select the optimal endogenous controls for ΔCt computation.

(h) Select groups of samples to compare and obtain relative quantification values.

4. It is highly recommended to test multiple endogenous controls for same type of your interest. Use geNorm or other analysis tools to identify better endogenous controls for data normalization. Based on published data and our experiences, we recommend U6, RNU24, RNU48, RNU44, U47, and

RNU6B for human samples and snoRNA-202, snoRNA-234, and snoRNA-420 for mouse samples (https://products.appliedbiosystems.com).

Acknowledgments

The authors would like to thank Dr. Neil Straus for comments and feedback on the manuscript.

References

1. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843–54.
2. Wightman, B., Ha, I., and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–62.
3. Ruvkun, G. (2001) Molecular biology. Glimpses of a tiny RNA world. *Science* **294**, 797–9.
4. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–97.
5. Griffiths-Jones, S., Saini, H.K., van Dongen, S., and Enright, A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**, D154–8.
6. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853–8.
7. Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858–62.
8. Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W. (2005) Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309**, 1573–6.
9. Orom, U.A., Nielsen, F.C., and Lund, A.H. (2008) MicroRNA-10a binds the 5′-UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell* 30, 460–71.
10. Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20.
11. Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005) Principles of microRNA/target recognition. *PLoS Biol* **3**, e85.
12. Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., and Kellis, M. (2005) Systematic discovery of regulatory motifs in human promoters and 3′UTRs by comparison of several mammals. *Nature* **434**, 338–45.
13. Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–73.
14. Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75–9.
15. Makeyev, E.V. and Maniatis, T. (2008) Multilevel regulation of gene expression by microRNAs. *Science* **319**, 1789–90.
16. Zhao, Y. and Srivastava, D. (2007) A developmental view of microRNA function. *Trends Biochem Sci* **32**, 189–97.
17. Visvanathan, J., Lee, S., Lee, B., Lee, J.W., and Lee, S.K. (2007) The microRNA miR-124 antagonizes the **...** pathway during embryonic CNS development. *Genes Dev* **21**, 744–9.
18. Conaco, C., Otto, S., Han, J.J., and Mandel, G. (2006) Reciprocal actions of REST and a microRNA promote neuronal identity. *Proc Natl Acad Sci USA* **103**, 2422–7.
19. Makeyev, E.V., Zhang, J., Carrasco, M.A., and Maniatis, T. (2007) The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol Cell* **27**, 435–48.

20. Boutz, P.L., Chawla, G., Stoilov, P., and Black, D.L. (2007) MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. *Genes Dev* **21**, 71–84.
21. Edwards, R.H., Marquitz, A.R., and Raab-Traub, N. (2008) Epstein-Barr virus BART miRNAs are produced from a large intron prior to splicing. *J Virol* **82**, 9094–106.
22. Zhang, R., Wang, Y.Q., and Su, B. (2008) Molecular evolution of a primate-specific microRNA family. *Mol Biol Evol* **25**, 1493–502.
23. Altuvia, Y., Landgraf, P., Lithwick, G., Elefant, N., Pfeffer, S., Aravin, A., Brownstein, M.J., Tuschl, T., and Margalit, H. (2005) Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res* **33**, 2697–706.
24. Gaur, A., Jewell, D.A., Liang, Y., Ridzon, D., Moore, J.H., Chen, C., Ambros, V.R., and Israel, M.A. (2007) Characterization of microRNA expression levels and their biological correlates in human cancer cell lines. *Cancer Res* **67**, 2456–68.
25. Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M., et al. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–14.
26. Rosenfeld, N., Aharonov, R., Meiri, E., Rosenwald, S., Spector, Y., Zepeniuk, M., Benjamin, H., Shabes, N., Tabak, S., Levy, A., et al. (2008) MicroRNAs accurately identify cancer tissue origin. *Nat Biotechnol* **26**, 462–9.
27. Chen, C., Ridzon, D., Lee, C.T., Blake, J., Sun, Y., and Strauss, W.M. (2007) Defining embryonic stem cell identity using differentiation-related microRNAs and their potential targets. *Mamm Genome* **18**, 316–27.
28. Xi, Y., Nakajima, G., Gavin, E., Morris, C.G., Kudo, K., Hayashi, K., and Ju, J. (2007) Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. *RNA* **13**, 1668–74.
29. Mitchell, P.S., Parkin, R.K., Kroh, E.M., Fritz, B.R., Wyman, S.K., Pogosova-Agadjanyan, E.L., Peterson, A., Noteboom, J., O'Briant, K.C., Allen, A., et al. (2008) Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci USA* **105**, 10513–18.
30. Chen, C., Ridzon, D.A., Broomer, A.J., Zhou, Z., Lee, D.H., Nguyen, J.T., Barbisin, M., Xu, N.L., Mahuvakar, V.R., Andersen, M.R., et al. (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res* **33**, e179.
31. Gilad, S., Meiri, E., Yogev, Y., Benjamin, S., Lebanony, D., Yerushalmi, N., Benjamin, H., Kushnir, M., Cholakh, H., Melamed, N., et al. (2008) Serum microRNAs are promising novel biomarkers. *PLoS One* **3**, e3148.
32. Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., et al. (2005) MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–8.
33. Calin, G.A., Liu, C.G., Sevignani, C., Ferracin, M., Felli, N., Dumitru, C.D., Shimizu, M., Cimmino, A., Zupo, S., Dono, M., et al. (2004) MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc Natl Acad Sci USA* **101**, 11755–60.
34. Krichevsky, A.M., King, K.S., Donahue, C.P., Khrapko, K., and Kosik, K.S. (2003) A microRNA array reveals extensive regulation of microRNAs during brain development. *RNA* **9**, 1274–81.
35. Nelson, P.T., Baldwin, D.A., Scearce, L.M., Oberholtzer, J.C., Tobias, J.W., and Mourelatos, Z. (2004) Microarray-based, high-throughput gene expression profiling of microRNAs. *Nat Methods* **1**, 155–61.
36. Sarkar, D., Parkin, R., Wyman, S., Bendoraite, A., Sather, C., Delrow, J., Godwin, A.K., Drescher, C., Huber, W., Gentleman, R., et al. (2008) Quality assessment and data analysis for microRNA expression arrays. *Nucleic Acids Res* **37**, e17.
37. Raymond, C.K., Roberts, B.S., Garrett-Engele, P., Lim, L.P., and Johnson, J.M. (2005) Simple, quantitative primer-extension PCR assay for direct monitoring of microRNAs and short-interfering RNAs. *RNA* **11**, 1737–44.
38. Shi, R. and Chiang, V.L. (2005) Facile means for quantifying microRNA expression by real-time PCR. *Biotechniques* **39**, 519–25.
39. Sharbati-Tehrani, S., Kutz-Lohroff, B., Bergbauer, R., Scholven, J., and Einspanier, R. (2008) miR-Q: a novel quantitative RT-PCR approach for the expression profiling of small RNA molecules such as miRNAs in a complex sample. *BMC Mol Biol* **9**, 34.
40. Chen, J., Lozach, J., Garcia, E.W., Barnes, B., Luo, S., Mikoulitch, I., Zhou, L., Schroth, G., and Fan, J.B. (2008) Highly sensitive and specific microRNA expression profiling using BeadArray technology. *Nucleic Acids Res* **36**, e87.

41. Allawi, H.T., Dahlberg, J.E., Olson, S., Lund, E., Olson, M., Ma, W.-P., Takova, T., Neri, B.P., and Lyamichev, V.I. (2004) Quantitation of microRNAs using a modified Invader assay. *RNA* **10**, 1153–61.
42. Jonstrup, S.P., Koch, J., and Kjems, J. (2006) A microRNA detection system based on padlock probes and rolling circle amplification. *RNA* **12**, 1747–52.
43. Schmittgen, T.D., Lee, E.J., Jiang, J., Sarkar, A., Yang, L., Elton, T.S., and Chen, C. (2008) Real-time PCR quantification of precursor and mature microRNA. *Methods* **44**, 31–8.
44. Mestdagh, P., Feys, T., Bernard, N., Guenther, S., Chen, C., Speleman, F., and Vandesompele, J. (2008) High-throughput stem-loop RT-qPCR miRNA expression profiling using minute amounts of input RNA**.** *Nucleic Acids Res* **36**, e143.
45. de Kok, J.B., Roelofs, R.W., Giesendorf, B.A., Pennings, J.L., Waas, E.T., Feuth, T., Swinkels, D.W., and Span, P.N. (2005) Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. *Lab Invest* **85**, 154–9.
46. Suzuki, T., Higgins, P.J., and Crawford, D.R. (2000) Control selection for RNA quantitation. *Biotechniques* **29**, 332–7.
47. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, RESEARCH0034.
48. Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**, 14863–8.
49. Andersen, C.L., Jensen, J.L., and Orntoft, T.F. (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* **64**, 5245–50.

Chapter 9

High-Throughput SuperSAGE

Hideo Matsumura, Kentaro Yoshida, Shujun Luo, Detlev H. Krüger, Günter Kahl, Gary P. Schroth, and Ryohei Terauchi

Abstract

SuperSAGE is a method of digital gene expression profiling that allows isolation of 26-bp tag fragments from expressed transcripts. Combined with the ultrahigh-throughput sequencing technologies, SuperSAGE enables analysis of millions of transcripts with lower cost and reduced effort and time. In this chapter, we present an updated protocol for this High-throughput SuperSAGE method with a special emphasis on the technique of library multiplexing.

Key words: Gene expression profiling, Transcriptome analysis, Next-generation sequencing, Multiplexing

1. Introduction

The rapid evolution of next-generation DNA sequencing technologies has considerably changed the way we study biological problems in various fields (1, 2). Massively parallel short-read sequencing technologies, so-called "Next Generation Sequencing" (NGS), allow whole genome de novo sequencing and resequencing at greatly reduced time, costs, and efforts. Another important application of these innovative sequencing technologies is gene expression (transcriptome) analysis. Sequencing-based gene expression analyses include Expressed Sequence Tag (EST) analysis (3), Serial Analysis Gene Expression (SAGE) (4), LongSAGE (5), SuperSAGE (6), Massively Parallel Sequencing Signatures (MPSS) (7), and RNA-Seq (8). Since current sequencing technologies allow the reading of billions of short DNA fragments without plasmid cloning, they are most suitable for short sequence tag-based expression analysis like SAGE or its derivatives. Apart

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_9, © Springer Science+Business Media, LLC 2011

from EST analysis, SuperSAGE obtains the longest tags (26–27 bp) from cDNAs. SuperSAGE tag sequences can clearly discriminate expressed genes from different organisms like host and pathogens (9), and therefore, allow researchers to analyze biological interactions among eukaryotes. Furthermore, longer cDNA fragments can easily be recovered by 3′-RACE PCR using complementary primers to the SuperSAGE tag sequences in nonmodel organisms, for which no EST or genome sequences are available (6). A combination of SuperSAGE and next-generation sequencing technology provides a powerful tool for various biological studies. At the moment, however, microarrays are still being employed as the dominant gene expression profiling tools. The major advantage of microarrays as compared to SuperSAGE (or other digital gene expression analyses) is the higher sample throughput (i.e., the number of different samples that can simultaneously be analyzed). However, most biological studies require analysis of gene expression of multiple samples as biological replicates or for kinetic studies.

We have established and present a protocol of the SuperSAGE technology combined with NGS and suitable for analyzing multiple samples, coined "High-Throughput (HT) SuperSAGE" using the Illumina Genome Analyzer, the machine for one of the Next-Generation Sequencing technologies (www.illumina.com). In this protocol, index (barcode) sequences are employed to discriminate tags from different samples (Fig. 1). Such barcodes enable researchers to analyze tags from many transcriptome samples in a single sequencing run by simply pooling the libraries.

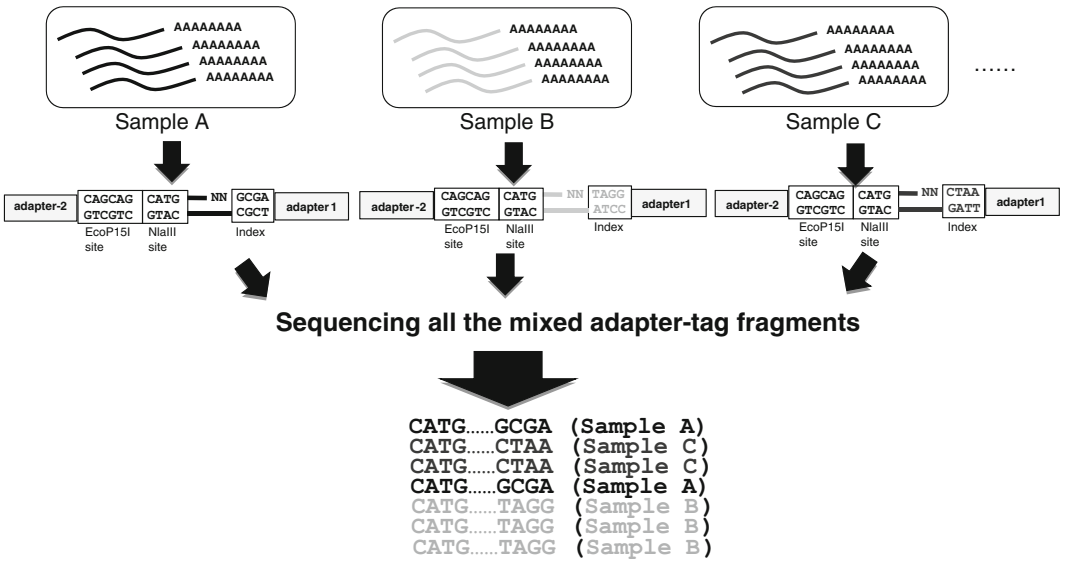


Fig. 1. Multiplexing strategy for sequencing analysis of SuperSAGE tags from several different samples. Adapter-1 with various index sequences is prepared and allocated to tags from different samples. Adapter-ligated tag fragments from several different samples are mixed and subjected to Next Generation Sequencing using the Illumina Genome Analyzer II. The resulting reads are separated by index sequences, and the sample origin of tags identified.

2. Materials

2.1. Adapter Preparation

1. Adapter oligonucleotides: adapter oligonucleotide synthesis and end-labeling are done by Operon Biotechnologies.

   Adapter-1sense:5′-ACAGGTTCAGAGTTCTACAG-TCCGACGATCXXXX-3′
   Adapter-1antisense: 5′-NNYYYYGATCGTCGGACTGTAG AACTCTGAACCTGT-amino-3′
   (XXXX encodes variable index sequences and should be complementary to YYYY, see Notes 1 and 2).
   Adapter-2sense: 5′-CAAGCAGAAGACGGCATACGATCTA ACGATGTACGCAGCAGCATG-3′
   Adapter-2antisense: 5′-CTGCTGCGTACATCGTTAGATC GTATGCCGTCTTCTGCTTG-amino-3′ (see Note 3)
   These oligonucleotides are purified with an Oligonucleotide Purification Cartridge (OPC) (see Note 4).
2. LoTE buffer: 3 mM Tris–HCl, pH 7.5, 0.2 mM EDTA.
3. Polynucleotide kinase buffer (10×, Takara): 0.5 M Tris–HCl, pH 8.0, 0.1 M $MgCl_2$, 50 mM DTT.

2.2. cDNA Synthesis

1. First strand buffer (5×, Invitrogen, Carlsbad, CA): 250 mM Tris–HCl, pH 8.0, 375 mM KCl, 15 mM $MgCl_2$.
2. Biotinylated adapter-oligo (dT) primer: Synthesized biotin-labeled oligonucleotides (5′-biotin-CTGATCTAGAGGTAC-CGGATCCCAGCAGTTTTTTTTTTTTTTTT-3′, HPLC-purified, Operon Biotechnologies, see Note 5) are dissolved in LoTE (1 μg/μL).
3. 0.1 M DTT (dithiothreitol, Invitrogen).
4. 10 mM dNTP (Invitrogen): 10 mM each of dATP, dTTP, dCTP, and dGTP.
5. SuperScript II reverse transcriptase (Invitrogen).
6. Second-strand buffer (Invitrogen): 100 mM Tris–HCl (pH 6.9), 450 mM KCl, 23 mM $MgCl_2$, 0.75 mM β-NAD^+, 50 mM $(NH_4)_2SO_4$.
7. *E. coli* DNA polymerase (10 U/μL, Invitrogen).
8. *E. coli* DNA ligase (1.2 U/μL, Invitrogen).
9. *E. coli* RNase H (2 U/μL, Invitrogen).
10. Binding buffer (PB buffer) in Qiaquick PCR purification kit (Qiagen).
11. Qiaquick spin column in Qiaquick PCR purification kit (Qiagen).
12. Washing buffer (PE buffer, 5×) (Qiagen): prepare 1× solution by adding ethanol before use.

2.3. Tag Extraction

1. NlaIII (10 U/μL, New England Biolabs): Store at –70°C.
2. NlaIII digestion buffer (NEBuffer 4) (10×, New England Biolabs): 20 mM Tris-acetate, pH 7.9, 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT.
3. BSA (10 mg/mL, New England Biolabs).
4. Streptavidin-coated magnetic beads (Dynabeads M-270 Streptavidin, see Note 6) (10 mg/mL, Invitrogen): Store at 4°C.
5. Siliconized microtubes (1.5 mL , Watson).
6. Binding and washing buffer (B&W buffer) (2×): 10 mM Tris–HCl, pH 7.5, 1 mM EDTA, 2 M NaCl.
7. T4 DNA ligase (2,000 U/μL, New England Biolabs): Store at –20°C.
8. T4 DNA ligase buffer (5×, New England Biolabs): 250 mM Tris–HCl, pH 7.5, 50 mM $MgCl_2$, 5 mM ATP, 50 mM DTT, 125 μg/mL BSA.
9. EcoP15I (10,000 U/μL, New England Biolabs, see Note 7): Store at –20°C.
10. EcoP15I digestion buffer (NEbuffer3) (10×, New England Biolabs): 500 mM Tris–HCl, pH 7.9, 1 M NaCl, 100 mM $MgCl_2$.
11. 10× ATP solution (1 mM, New England Biolabs).
12. 100× BSA solution (100 μg/mL, New England Biolabs).
13. Phenol: chloroform: isoamyl alcohol (25:24:1) (Invitrogen). Store at 4°C.
14. Ammonium acetate: 10 M solution (Wako Chemicals).
15. Glycogen solution (20 mg/mL, Roche Diagnostics).

2.4. Indexed-Adapter Ligation and PCR

1. Phusion HF 5× Buffer (Finnzymes).
2. dNTP solution (Invitrogen): 10 mM each of dATP, dTTP, dCTP, and dGTP.
3. $MgCl_2$ solution (50 mM, Finnzymes).
4. Adapter-1 primer: 5′-AATGATACGGCGACCACCGACAGG TTCAGAGTTCTACAGTCCGA-3′.

   Adapter-2 primer: 5′-CAAGCAGAAGACGGCATACGA-3′.

   These oligonucleotides (see Note 8) are synthesized and purified via an Oligonucleotide Purification Cartridge (OPC) procedure by Operon Biotechnologies. They are dissolved in LoTE at 100 pmol/μL.
5. Phusion Hot Start DNA polymerase (2 U/μL, Finnzymes): Store at –20°C.
6. Acrylamide/Bis solution (40%, 19:1, Bio-Rad): Store at 4°C.

7. *N*,*N*,*N*,*N*′-Tetramethyl-ethylenediamine (TEMED, Wako Chemicals). Store at 4°C.
8. Ammonium persulfate: prepare 10% solution in sterilized water and store at 4°C.
9. 6× loading dye: 30% (v/v) glycerol, 0.25% (w/v) bromophenol blue and 0.25% (w/v) xylene cyanol.
10. SYBR green solution: Original SYBR green I stock solution (Molecular Probes) is diluted 10,000 times with 1× TAE buffer. Store at 4°C.
11. 20 bp DNA ladder marker (200 ng/μL, Takara).

2.5. Purification of PCR Product

1. ERC buffer in MinElute Reaction Cleanup kit (Qiagen, Germany).
2. MinElute spin column in MinElute Reaction Cleanup kit (Qiagen, Germany).
3. Spin-X column (Corning, Corning, NY).

2.6. Multiplexing DNA Samples for Sequencing Analysis

1. Agilent 2100 Bioanalyzer (Agilent Technologies, Wilmington, DE).
2. Agilent DNA 1000 kit includes chips and Gel-Dye mix (Agilent Technologies, Wilmington, DE).

3. Methods

In this High-Throughput SuperSAGE method using the Illumina Genome Analyzer II (GAII), the procedures for cDNA synthesis and 26-bp tag extraction using EcoP15I are almost identical to the original SuperSAGE protocol (6, 8), except that ditags are no longer formed before cloning and sequencing. Instead, a single tag is ligated to two different adapters. This single tag-based analysis simplifies the protocol and takes advantage of the millions of short, high-quality read that can be generated on the Genome Analyzer II system. Increased read length in the Illumina Genome Analyzer IIx (upto 100 bp) allows analysis of ditags; however, this single tag protocol is quicker and easier to perform than the ditag method, uses less PCR amplification, and gives highly reproducible transcript profiling data. For multiplexing SuperSAGE tag-libraries in a single sequencing run, adapter fragments harboring different index sequences are ligated to the tags derived from different biological samples (Fig. 1). After sequencing of pooled indexed fragments (adapter-tags), the sequence reads obtained are separated by their index sequences. Employing this procedure, we have successfully analyzed the transcriptomes of multiple samples in a single lane on the GAII.

3.1. Adapter Preparation

1. Dissolve adapter oligonucleotides (adapter-1 sense, adapter-1 antisense, adapter-2 sense, adapter-2 antisense) in LoTE buffer (100 pmol/μL).
2. Mix 10 μL of each sense and antisense complementary oligonucleotide solution of adapter-1 or adapter-2 in each tube. Add 3 μL 10× polynucleotide kinase buffer and 7 μL LoTE.
3. A total of 30 μL of this mixture is denatured at 95°C for 2 min and cooled down to 20°C for annealing complementary oligonucleotides. The annealed double-stranded DNAs are designated adaper-1 and adapter-2, respectively.

3.2. cDNA Synthesis

1. The synthesis of double-stranded cDNA follows the protocol of the SuperScriptII double-strand cDNA synthesis kit (Invitrogen). Total RNA (2–10 μg) is dissolved in 11 μL DEPC-treated water and incubated at 70°C for 10 min after adding 1 μL biotinylated adapter-oligo dT primer (100 pmol). Denatured RNA solution is immediately placed on ice, and 4 μL 5× First Strand buffer, 2 μL 0.1 M DTT, 1 μL 10 mM dNTP, and 1 μL SuperScriptII reverse transcriptase are added for first strand cDNA synthesis. The reaction mixture is incubated at 45°C for 1 h.
2. For second strand cDNA synthesis, 30 μL 5× Second Strand buffer, 91 μL sterile water, 3 μL 10 mM dNTP, 4 μL *E. coli* DNA polymerase, 1 μL *E. coli* RNase H, and 1 μL *E. coli* DNA ligase are added to 20 μL first strand cDNA solution and mixed. Incubate at 16°C for 2 h.
3. To purify the double-stranded cDNA, 750 μL of PBI buffer from Qiaquick PCR purification kit (Qiagen) are added. The mixed solution is then applied to a Qiaquick spin column and centrifuged at 10,000 × *g* for 1 min. After discarding the flow-through, 750 μL washing buffer (PE buffer) is applied to the column. Centrifuge at 10,000 × *g* for 1 min and discard flow-through. For complete drying, the column is centrifuged at maximum speed for 1 min. After the column is transferred to a new 1.5-mL microtube, 30 μL LoTE is added for elution. The eluant (purified cDNA) is collected by centrifugation at 10,000 × *g* for 1 min.

3.3. Tag Extraction

1. For digestion of cDNA, 20 μL NlaIII digestion buffer (NEBuffer 4), 2 μL BSA, 152 μL LoTE, and 5 μL NlaIII are added to the cDNA solution, mixed, and incubated at 37°C for 1.5 h.
2. Prepare 100 μL suspension of streptavidin-coated magnetic beads (Dynabeads M-270) in a siliconized 1.5-mL microtube (see Note 9). Place the tubes containing magnetic beads on

a magnetic stand and remove the supernatant with a pipette. For washing the magnetic beads, 200 μL 1× B&W solution is added, and beads are suspended well by pipetting. Place the tube on a magnetic stand, and remove and discard the supernatant.

3. To the washed magnetic beads, 200 μL 2× B&W solution and 200 μL digested cDNA solution are added and suspended well. Leave the tube for 15–20 min at room temperature with occasional mixing, so that the biotinylated cDNAs readily bind to streptavidin on the magnetic beads. After digested cDNAs have associated with the beads, the tube is placed on the magnetic stand, and the supernatant is discarded. The magnetic beads are washed three times with 200 μL 1× B&W and once with 200 μL LoTE.
4. For adapter-2 ligation to the digested cDNAs bound on the magnetic beads, 21 μL LoTE, 6 μL 5× T4 DNA ligase buffer, and 1 μL adapter-2 solution are added to the magnetic beads. After mixing with a pipette, the bead suspension is incubated at 50°C for 2 min to dissociate adapter dimers. Tubes are kept at room temperature for 15 min. After the tubes have cooled down, 2 μL T4 DNA ligase (10 U) is added and incubated at 16°C for 2 h with occasional mixing.
5. After the ligation reaction, beads are washed four times with 1× B&W, and three times with LoTE. The beads are suspended in 75 μL LoTE.
6. For EcoP15I digestion of the fragments on the magnetic beads, 10 μL 10× NEBuffer 3, 10 μL 10× ATP solution, 1 μL 100× BSA, and 4 μL EcoP15I are added to the suspended magnetic beads. Incubate the tube at 37°C for 2 h with occasional mixing (see Note 10).
7. Following EcoP15I digestion, the bead suspension is placed on the magnetic stand and the supernatant collected into a new tube. The beads are resuspended in 100 μL 1× B&W. After separation on the magnetic stand, the supernatants are retrieved and combined to the previously collected solution (see Note 11).
8. To the collected solution, containing the adapter-tag fragments, half a volume of phenol: chloroform:isoamyl alcohol (195 μL) is added, shortly vortexed and spun at 10,000×*g* for a few minutes. The upper aqueous layer is transferred to a new tube. For ethanol precipitation, 100 μL 10 M ammonium acetate, 3 μL glycogen, and 900 μL cold ethanol are added to the collected solution (approximately 200 μL). Keep the tube at –80°C for 1 h and then centrifuge at maximum speed for 40 min at 4°C. The resulting pellet is washed twice with 70% (v/v) ethanol and dried. Precipitated adapter-2 ligated 26 bp-tag fragments are dissolved in 10 μL LoTE.

3.4. Indexed-Adapter Ligation and PCR

1. Prepare adapter-1 with defined index sequences assigned to individual samples (adapter-1a, 1b, 1c…). For ligation of adapter-1, 3 μL 5× T4 DNA ligase buffer and 0.5 μL adapter-1 solution are added to the adapter-2 ligated tags. Incubate the tube at 50°C for 2 min and keep it at room temperature for 15 min. After the tubes have cooled down, 1.5 μL T4 DNA ligase (7.5 U) is added and incubated at 16°C for 2 h (see Note 12).
2. For PCR amplification of adapter-ligated tag fragments, a reaction mixture containing 3 μL 5× Phusion HF buffer, 0.3 μL 2.5 mM dNTP, 0.1 μL 50 mM $MgCl_2$, 0.15 μL adapter-1 primer, 0.15 μL adapter-2 primer, 10.1 μL distilled water, 1 μL ligation solution, and 0.2 μL Phusion Hot Start DNA polymerase is prepared in a tube (see Note 13).
3. PCR cycle: 98°C for 2 min, then 5–10 cycles each at 98°C for 30 s, and 60°C for 30 s (see Note 14).
4. The size of the amplified PCR product is confirmed by polyacrylamide gel electrophoresis (PAGE, see Note 15). Prepare an 8% PAGE gel by mixing 3.5 mL 40% acrylamide/bis solution, 13.5 mL distilled water, 350 μL 50× TAE buffer, 175 μL 10% ammonium persulfate, and 15 μL TEMED. Pour the solution onto the gel plate (12 cm × 12 cm, 1 mm thickness), and insert a comb (no stacking gel).
5. Running buffer (1× TAE) is prepared and added to both the upper and lower electrophoresis chambers. Then, 3 μL 6× loading dye is added to 15 μL of the PCR solution and the mixture loaded to the well. Two μL of a 20 bp molecular size marker is also loaded. Run the gel at 75 V for 10 min and then at 150 V for around 30 min (until the BPB dye front has migrated two-thirds down the gel).
6. After electrophoresis, the gel is removed from the plate. Pour 1 mL SYBR green solution (diluted in 1× TAE buffer) on the plastic wrap and place the gel on it. Further, disperse 1 mL SYBR green solution onto the gel. After a 2 min staining period, the gel is placed on a UV transilluminator. The size of the expected amplified fragment (tag sandwiched with two adapters) is 123–125 bp (Fig. 2, see Note 16).
7. After the confirmation of PCR amplification of the adapter-ligated tag fragments, repeat PCRs at the same condition using 14 tubes.

3.5. Purification of PCR Product

1. After PCR, solutions from all tubes are collected into a 1.5-mL tube and 400 μL of ERC buffer (attached to a MinElute Reaction Cleanup kit) is added. Prepare a MinElute spin column from the same kit and transfer a mixture of PCR solution and ERC buffer onto the column. Centrifuge at 10,000 × *g* for

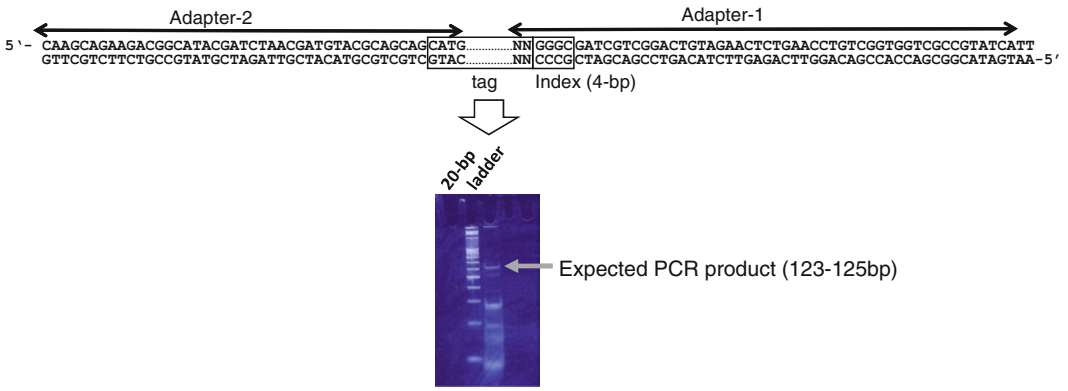


Fig. 2. Structure of the PCR amplified adapter-ligated tag. The regions of adapter-1, adapter-2, and tag sequences are indicated by *arrows*. The index sequence in the adapter-1 is *boxed*. The gel picture shows the result of an 8% PAGE run of PCR products. The band of the expected PCR product (adapter-ligated tag; 123–125 bp) is indicated by an *arrow*. Only the band of this size is excised from the gel and purified for sequencing. Smaller DNA bands (55 and 36 bp) in the gel are derived from unligated adapter-1 (see Note 21). *Left lane:* 20-bp size marker ladder.

1 min and discard flow-through. Add another 750 μL of washing buffer (PE buffer) to the column. Centrifuge at 10,000 × *g* for 1 min and discard flow-through. For complete drying of the columns, centrifuge at maximum speed for 1 min. After columns are placed on a new 1.5-mL microtube, 15 μL LoTE is added to the column for elution. Leave the column for 1 min after adding LoTE, centrifuge at 10,000 × *g* for 1 min and collect eluate.

2. Prepare 8% polyacrylamide gel as described in Subheading 3.4. Add 3 μL 6× loading buffer to column-purified PCR product and load it into the well (see Note 17). After running the gel as described in Subheading 3.4, it is stained with SYBR green, and bands are visualized under UV light.
3. Only the 123–125 bp band (26–27 bp tag flanked by adapter-1 and adapter-2) is cut out from the gel and transferred to a 0.5-mL microtube (see Note 18). Make holes at the top and the bottom of the tube with a needle and place it in a 1.5-mL tube. Centrifuge the tube at maximum speed for 2–3 min. Polyacrylamide gel pieces sediment at the bottom of the tube. Add 300 μL LoTE to the gel pieces and suspend.
4. After incubation at 37°C for 2 h, the gel suspension is transferred to a Spin-X column and centrifuged at maximum speed for 2 min. The eluate is once extracted by phenol/chloroform and precipitated by adding 100 μL 10 M ammonium acetate, 3 μL glycogen, and 950 μL cold ethanol. Keep it at −80°C for 1 h and then centrifuge at 15,000 × *g* for 40 min at 4°C. Wash once with 70% ethanol and dry. The resulting pellet is then dissolved in 10–15 μL LoTE.

3.6. Multiplexing DNA Samples for Sequencing Analysis

1. For quantifying the purified PCR product, it is analyzed on an Agilent Bioanalyzer system (see Note 19). A DNA chip from Agilent DNA 1000 kit is prepared and filled with Gel-Dye Mix supplied with the kit. Load 1 μL purified PCR product into the well of the chip, and run the chip on the Agilent 2100 Bioanalyzer.
2. The DNA concentration of the 123–125 bp fragment is measured using 2100 Expert software (Agilent Technologies). Based on this quantification, equal amount of DNA (PCR product) from each sample is mixed and the pooled mixture (10 μL of 10 nM, see Note 20) is applied to sequencing analysis by the Illumina Genome Analyzer II.

3.7. Sequence Data Analysis

1. For 26 bp-tag extraction from sequence reads analyzed by the Genome Analyzer II and counting of tags, we made our own programs written in Perl language (SuperSAGE_data_parse1) which are available upon request.
2. First, the sequence reads are separated into independent files according to the first 4-base index sequences in each read. From separated sequence reads in each file, extract 22-base sequences upstream of 5′-CATG-3′ (NlaIII site).
3. Subsequently, count redundant tag sequences in each file. A list of independent tag sequences and their counts is obtained.

4. Notes

1. Adapter-1 contains a complementary site for the sequencing primer (5′-CCGACAGGTTCAGAGTTCTACAGTCCGA CATG-3′), followed by the corresponding index sequence for the discrimination of different samples. Our current protocol employs 4-bp index sequences.
2. To facilitate ligation to tags, the end of adapter-1 has a 2-base 5′-protrusion, compatible with the end of the EcoP15I-digested fragment. Also, its digested ends carry various 2-base protrusions. Thus, 2-bases of 5′-end in adapter-1 antisense oligonucleotides are synthesized as “NN” (mixture of any nucleotides).
3. Adapter-2 contains a recognition site for EcoP15I, and its end should be compatible with the end of an NlaIII-digested fragment.
4. To prevent incorrect ligation of adapters to tags, the 3′-ends of adapter-1 and adapter-2 antisense oligonucleotides are amino-modified.
5. An EcoP15I-recognition site (5′-CAGCAG-3′) is contained within the biotinylated adapter-oligo dT primer.

6. Dynabeads M-270 Streptavidin represent beads with a hydrophilic surface, which shows less binding of nonspecific DNA fragments.
7. EcoP15I, an enzyme with high specific activity, is now commercially available, while our own enzyme was only used in the original SuperSAGE protocol (6).
8. Adapter-1 and adapter-2 primers are identical to primers of the Digital Gene Expression analysis kit from Illumina.
9. Employing beads with hydrophilic surfaces and siliconized microtubes, most of unligated adapters can be eliminated by washing with buffers.
10. Within the adapter-2 ligated cDNA fragments bound to magnetic beads, two EcoP15I sites are present. The enzyme sometimes recognizes the site adjacent to the 3′-poly-A, leading to the release of fragments longer than the adapter-tag after digestion.
11. By repeated washing of the paramagnetic beads, residual adapter-tag fragments can be collected.
12. Tubes can be incubated for longer periods of time (e.g., overnight).
13. When preparing the PCR mixture, care should be taken so that no contamination with previously amplified PCR products occurs. Use pipettes and solutions, including water, separate from those used in the experiments after PCR. Also, use separate labware and gloves.
14. PCR amplification of each sample should last 10 or more cycles, since it is most important to securely confirm the amplification of the adapter-ligated tag at this step.
15. PAGE is better than agarose gel electrophoresis to clearly separate small fragments (<100 bp).
16. PCR products derived from long fragments (as described in Note 14) are observed as smear larger than 125 bp in the gel.
17. Loading too much DNA into the well reduces resolution of PAGE. Normally, purified PCR products from 14 reactions are loaded in two wells separately.
18. Any contamination with DNA fragments of inappropriate size (larger or smaller than 123–125 bp) reduce the number of reads containing tag sequences, since any DNA with adapter sequences are equally processed and sequenced by the GAII.
19. It is quite important to accurately measure the DNA amount of samples in orderto obtain the expected number of sequencing reads in each sample after multiplexing. For quantification of DNA, either the Bioanalyzer or the fluorescent (SYBR green I) method is recommended rather than spectrophotometric measurement.

20. In the case of the 123–125 bp PCR product, a total of 10 ng of pooled DNA is sufficient for GAII sequencing analysis.
21. The size of unligated adapter-1 is 36 bp, and the fragment at 55 bp is probably produced by extension and annealing of adapter-1 primer to unligated adapter-1.

Acknowledgments

HM is supported by the Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry (BRAIN).

References

1. Morozova, O. and Marra, M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**, 255–64.
2. Metzker, M.L. (2010) Sequencing technologies – the next generation. *Nat. Rev. Genet.* **11**, 31–46.
3. Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651–6.
4. Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995) Serial analysis of gene expression. *Science* **270**, 484–7.
5. Saha, S., Sparks, A.B., Rago, C., Akmaev, V., Wang, C.J., Vogelstein, B., Kinzler, K.W., and Velculescu, V.E. (2002) Using the transcriptome to annotate the genome. *Nat. Biotechnol.* **20**, 508–12.
6. Matsumura, H,, Reich, S., Ito, A., Saitoh, H., Kamoun, S., Winter, P., Kahl, G., Reuter, M., Kruger, D.H., and Terauchi, R. (2003) Gene expression analysis of plant host-pathogen interactions by SuperSAGE. *Proc. Natl. Acad. Sci. USA* **100**, 15718–23.
7. Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J., and Corcoran, K. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18**, 630–4.
8. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–8.
9. Matsumura, H., Ito, A., Saitoh, H., Winter, P., Kahl, G., Reuter, M., Kruger, D.H., and Terauchi, R. (2005) SuperSAGE. *Cell. Microbiol.* **7**, 11–8.

Chapter 10

Deep Cap Analysis of Gene Expression

Junpei Kurosawa, Hiromi Nishiyori, and Yoshihide Hayashizaki

Abstract

The cap analysis of gene expression (CAGE) technology has been established to detect transcriptional starting sites (TSSs) and expression levels by utilizing 5′ cDNA tags and PCR. It has been reported that the amount of templates is proportional to the amplification efficiency of PCR. CAGE has been used as a key technique for analyzing promoter activity and finding new transcripts including alternative spliced products and noncoding transcripts. Here, we introduce more powerful tools such as deepCAGE, which can be utilized for high-throughput next-generation sequencing technology. DeepCAGE can produce much deeper transcriptome datasets and can reveal more details of the regulatory network.

Key words: Transcriptome, Transcription start sites, Cap trap, DeepCAGE, High-throughput sequencer

1. Introduction

Genome sequencing and cDNA projects have revealed the complexity of transcriptomes by finding not only coding regions but also a large proportion of noncoding regions. In studies of the complexity of the transcriptomes in cells and tissues, expression profiling using microarray, qRT-PCR, or cap analysis of gene expression (CAGE) analysis (1) is crucial. CAGE technology has been playing an especially pivotal role in transcriptome analysis because it permits identification of transcription start sites (TSSs) with the associated core promoter regions (2, 3).

Here, we introduce a new version of the CAGE protocol that has been modified for applications with next-generation high-throughput sequencers: it is called "deepCAGE" (4) (Fig. 1). The deepCAGE protocol starts with cDNAs synthesis by using random oligo primers and biotinylation. The single-stranded RNA is removed with RNaseI and single cDNA strands are

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_10, © Springer Science+Business Media, LLC 2011

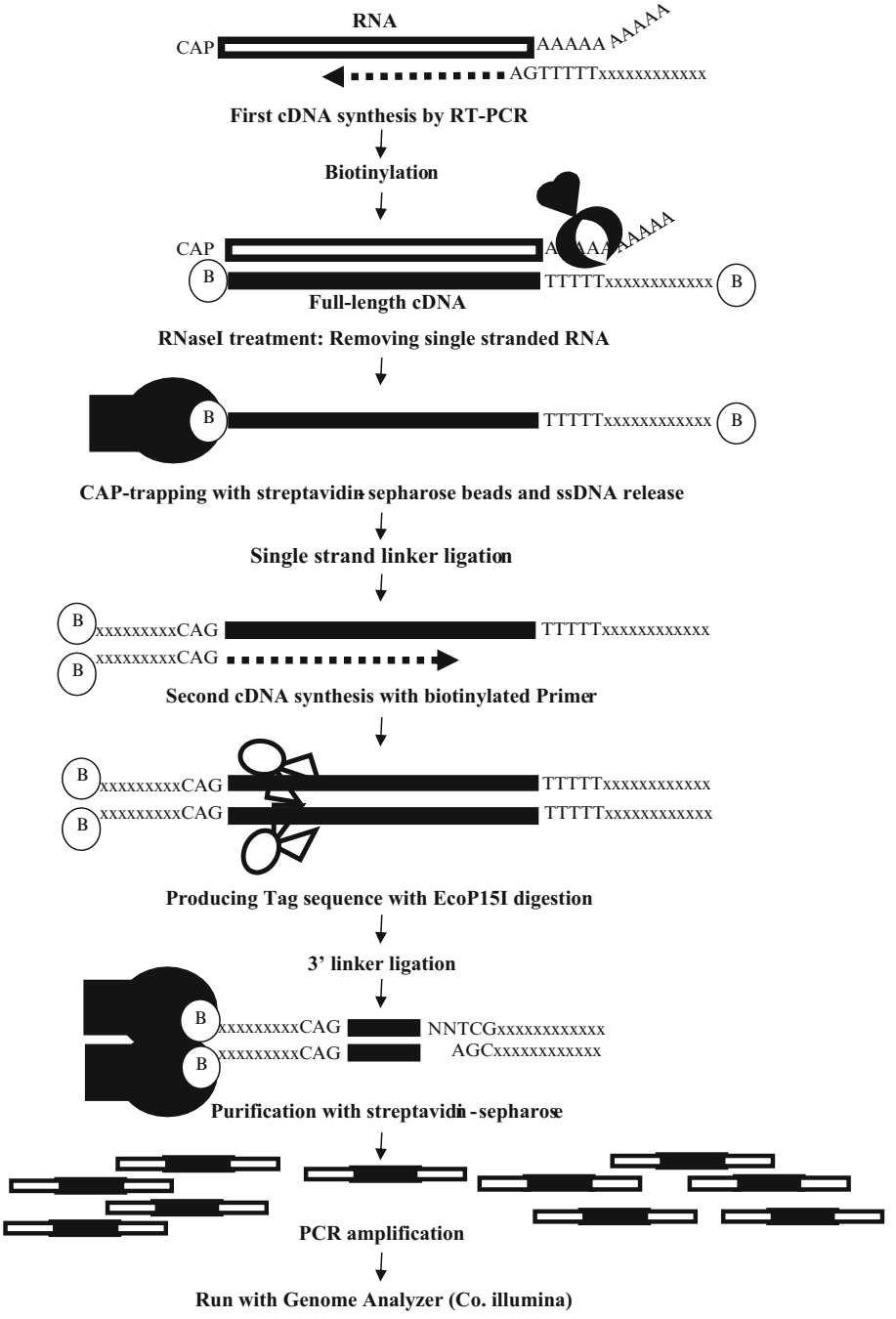


Fig. 1. A schematic diagram of the deepCAGE method: the deepCAGE protocol starts with cDNA synthesis by using random oligo primers and biotinylation. The single-stranded RNA is removed with RNaseI and single cDNA strands are cap-trapped using streptavidin-sepharose beads; then, 5′ linker, which has the biotin and EcoP15I recognition sequence, is ligated. After synthesis of the second cDNA strands, the products are digested with EcoP15I, which recognizes 5′ linker and cuts about 25 bases from the 5′ end to yield the tag sequence. Cleaved tag sequences are ligated to the linker at the 3′ end. After purification with streptavidin-sepharose beads from the 3′ linker, PCR that is proportional to the amount of template is performed. PCR products are deposited on recent shotgun types of sequencers.

cap-trapped using streptavidin-sepharose beads and then 5′ linker, which has the biotin and EcoP15I recognition sequence, is ligated. After synthesis of the second cDNA strands, the products are digested with EcoP15I, which recognizes 5′ linker and cuts about 25 bases from the 5′ end to yield the tag sequence. Cleaved tag sequences are ligated to the linker at the 3′ end. After purification with streptavidin-sepharose beads from the 3′ linker, PCR that is proportional to amount of templates is performed. PCR products are purified by PCI/CIA and EtOH precipitation and are deposited on recent shotgun-types of sequencers so that we can perform sequencing of whole redundant CAGE tags to measure the expression levels of all target loci.

2. Materials

2.1. Common, Reagents, Buffer, Equipment

1. RTase (200 U/μL, Promega; RNaseH minus, point mutant).
2. ProK (QIAGEN).
3. 80% (v/v) EtOH.
4. 99.5% (v/v) EtOH.
5. 0.5 M EDTA-NaOH (pH 8.0).
6. 1 M NaOAc-acetic acid (pH 4.5).
7. 250 mM $NaIO_4$ (53.47 μg/μL in water, ICNBiomedicals, Inc.: Sodium Periodate ACS Reagent Grade).
8. 40% (v/v) glycerol.
9. 1 M Sodium citrate–HCl (pH 6.0).
10. 1 M Tris–HCl (pH 8.5).
11. RNaseI (10 U/μL).
12. 10% SDS.
13. 5 M NaCl.
14. Streptavidin-Sepharose High Performance (GE Healthcare).
15. 50 mM NaOH.
16. 1 M Tris–HCl (pH 7.0).
17. TAKARA DNA ligation Kit ver.2.1 I (TAKARA).
18. TAKARA DNA ligation Kit ver.2.1 II (TAKARA).
19. GFX column (GE Healthcare).
20. MicroSpin Column (GE Healthcare).
21. LA Taq buffer (10×, TAKARA).
22. 25 mM $MgCl_2$.
23. 2.5 mM dNTPs.

24. LA Taq (5 U/μL, TAKARA).
25. NEbuffer3 (10×).
26. BSA (100×) (supplied with EcoP15I).
27. ATP (10×) (supplied with EcoP15I).
28. 10 mM Sinefungin (Merck Ltd.).
29. EcoP15I (1 U/μL, NEB).
30. 1 M $MgCl_2$.
31. T4 DNA ligase (400 U/μL, NEB).
32. HF buffer (5×).
33. Phusion High-Fidelity DNA polymerase (2 U/μL, FINNZYMES).
34. 12% (w/v) PAGE gel.
35. β-agaraseI (1×) buffer.
36. β-agaraseI (1 U/μL).
37. MicroconYM-100 (Millipore).
38. Glycogen (1 μg/μL).

2.2. Sorbitol/Trehalose

1. Trehalose (Fluka, anhydrous 72.4%) saturated solution: Add 7.27 g trehalose to water up to 10 mL and autoclave at 121°C for 30 min.
2. 4.9 M Sorbitol (Fluka) solution: Add 17.8 g sorbitol to water up to 20 mL and autoclave at 121°C for 30 min.
3. Mix the solutions trehalose to sorbitol = 1:2.
4. Transfer the solution to a 50-mL disposable tube.
5. Scrape up Chelex100 Resin with a disposable pipette and add about 1 cm to the solution (see Note 1).
6. Vortex and keep at room temperature (RT) for 3 h.
7. Transfer the solution to a 1.5-mL tube and centrifuge at 15,000 × *g* for several minutes. Collect the supernatant.

2.3. High SALT CTAB Solution

1. Mix urea (final concentration of 4 M), CTAB [final concentration of 1% (v/v)], 1 M Tris–HCl (pH 7.0) (final concentration of 50 mM), 0.5 M EDTA (pH 8.0) (final concentration of 1 mM), and water to make 1% CTAB-urea.
2. Mix 10 mL 1% CTAB-urea and 2.01 mL 5 M NaCl (final concentration of 0.84 M).

2.4. Wash Buffer

Mix 1 M Tris–HCl (pH 7.0) (final concentration of 100 mM), 0.5 M EDTA (pH 8.0) (final concentration of 5 mM), 5 M NaCl (final concentration of 0.3 M), 100% EtOH [final concentration of 60% (v/v)], and water.

2.5. Binding and Wash Buffer (B&W Buffer) Mix 0.5 M EDTA (pH 8.0) (final concentration of 50 mM), 5 M NaCl (final concentration of 0.5 M), and water.

2.6. B&W Buffer with 0.05% SDS Mix B&W buffer and 10% SDS [final concentration of 0.05% (v/v)].

2.7. B&W Buffer with 0.1% Tween 20 Mix B&W buffer and Tween 20 [final concentration of 0.1% (v/v)].

2.8. Buffer (5×) Mix 1 M Tris–HCl (pH 7.0) (final concentration of 250 mM), 100 mM ATP (final concentration of 5 mM), BSA (10 mg/mL) (final concentration of 25 μg/mL), and water.

2.9. 10 mM Biotin Hydrazide Add 3.71 μg of Biotin (long arm) Hydrazide (Vector Laboratories) to 150 μL of water.

2.10. 5% (w/v) Agarose TAE Gel

1. Weigh out 5 g of agarose and feed it into Erlenmeyer flask.
2. Pour 100 mL of TAE buffer (1×) into the flask.
3. Wrap in plastic wrap and dissolve the above reagents by heating in a microwave.
4. After complete dissolution, slowly stir the solution with a stirring bar until you could touch the flask by hand.
5. Pipette 1 μL of ethidium bromide (10 mg/mL) into the solution.
6. Pour the solution into a plate.

2.11. PCI

1. Mix phenol,chloroform, and isoamyl alcohol at 25:24:1.

2.12. CIA

1. Mix chloroform and isoamyl alcohol at 24:1.

2.13. Oligonucleotides

1. First-strand synthesis Random RT primer: 5′ AAGGTCTAT CAGCAGNNNNNNNNNNNNC 3′.
2. First-strand synthesis Oligo dt RT primer: 5′ AAGGTCTATC AGCAGTTTTTTTTTTTTTTVN 3′.
3. 5′ Linker Upper: 5′ Bio CCACCGACAGGTTCAGAGTTCT ACAGAGACAGCAGNNNNNN 3′

   Lower: Phos CTGCTGTCTCTGTAGAACTCTGAACCT GTCGGTGG NH_2 3′ (see Note 2).
4. 5′ primer: 5′ Bio CCACCGACAGGT TCAGAGTTCTACAG 3′.
5. 3′ Linker: Upper: 5′ Phos NNTCGTATGCCGTCTTCTGC TTG 3′

   Lower: 5′ CAAGCAGAAGACGGCATACGA 3′ (see Note 2).
6. PCR Forward primer: 5′ AATGATACGGCGACCACCGAC AGGTTCAGAGTTC 3′.

7. PCR Reverse primer: 5′CAAGCAGAAGACGGCATACGA 3′.

2.14. Sequencer

1. Genome Analyzer version; GA, GAII, or GAIIx (Illumina).

2.15. Analysis for Sequencing Result

1. GA pipeline (Illumina): This program for basecalling is supplied by Illumina.
2. Eland (Illumina): This program for mapping is supplied by Illumina.

3. Methods

3.1. First Strand

3.1.1. First-Strand Synthesis

1. Prepare total RNA from cells or tissue using standard methods, for example, TRIzol Reagent.
2. Mix 50 μg of total RNA and 12 μg of RT primer – mix 1 (see Note 3).
3. Incubate at 65°C for 10 min to denature and then cool on ice for 2 min.

Buffer (5×)	20 μL	
10 mM dNTPs	2.7 μL	0.27 mM
Sorbitol/trehalose	20 μL	20% (see Note 4)
RTase (200 U/μL)	15 μL	
Water	Total volume = 100 μL	Including RNA and primer

4. During the incubation, mix the above reagents – mix 2.
5. Cool on ice, then combine mix 1 (described in Subheading 3.1.1, step 2) and mix 2 (described in Subheading 3.1.1, step 4) cocktails and keep on ice.
6. Reverse transcription is performed on a thermal cycler with the following program: 25°C for 30 s – 42°C for 30 min – 50°C for 10 min – 56°C for 10 min – 60°C for 10 min – 4°C hold.

3.1.2. ProK Treatment

1. Add 2 μL of ProK (20 mg/mL) and 4 μL of 0.5 M EDTA (pH 8.0) for each 100 μL of reaction solution.
2. Incubate at 45°C for 20 min, then cool on ice for 2 min.

3.1.3. GFX-CTAB Purification

RT solution	106 μL
High SALT CTAB solution	160 μL

1. Mix all of these solutions and incubate at 65°C for 10 min. Next, leave at room temperature for 10 min to bind RNA/cDNA fragments to CTAB.

2. Pour the solution into the GFX column. Centrifuge at 4,400 × *g* for 1 min and discard the flow-through.
3. Pour 600 µL of wash buffer into the column. Centrifuge at 4,400 × *g* for 1 min and discard the flow-through.
4. Pour 600 µL of 80% EtOH into the column. Centrifuge at 15,000 × *g* for 1 min and discard the flow-through.
5. Place the column on a new 1.5 mL tube.
6. Add 85 µL of 65°C preincubated water and keep at room temperature for 1–2 min.
7. Centrifuge at 15,000 × *g* for 1 min (see Note 5).
8. Add 10 µL of 65°C preincubated water and centrifuge at 15,000 × *g* for 1 min.
9. Collect the eluent.

3.1.4. Biotinylation

RNA/cDNA solution	90 µL
1 M NaOAc (pH 4.5)	6.5 µL
250 mM $NaIO_4$	4 µL

1. Mix all of the above mentioned reagents (see Note 6).
2. Cool on ice for 45 min in dark.
3. To stop the reaction, add 2 µL of 40% glycerol and mix well by vortexing.
4. Add 400 µL of water to the above solution.
5. Pour the solution into the MicroconYM-100 and centrifuge at 600 × *g* for 25 min at room temperature. Discard the flow-through (see Note 7).
6. Add 300 µL of water again and centrifuge at 600 × *g* for 25 min at room temperature.
7. Recover the cDNA by inverting the filter in a new tube and centrifuge at 7,300 × *g* for 2 min.
8. Add appropriate volume of water to the filter column and centrifuge at 7,300 × *g* for 1–2 min (see Note 8).
9. Repeat step 7.

Recovered solution	50 µL (adjust to 50 µL with water)
1 M Sodium citrate (pH 6.0)	5 µL
10 mM Biotin (Long Arm) Hydrazide (see Note 9)	150 µL

10. Mix all of the above mentioned reagents .
11. Incubate at room temperature overnight.

3.1.5. RnaseI Treatment

Biotinylated solution	205 μL
1 M Tris–HCl (pH 8.5)	20 μL
0.5 M EDTA (pH 8.0)	4 μL
RNaseI (10 U/μL)	10 μL

1. Mix all of the above mentioned solutions and incubate at 37°C for 30 min to digest single-strand RNA.
2. Incubate at 65°C for 5 min to stop the reaction.
3. Add 3 μL of 10% SDS and 3 μL of ProK (20 mg/mL).
4. Incubate at 45°C for 15 min.
5. Add 250 μL of water to the above solution.
6. Pour the solution into MicroconYM-100 and centrifuge at 600×*g* for 25 min at room temperature. Discard the flow-through (see Note 7).
7. Add 400 μL of water again and centrifuge at 600×*g* for 25 min at room temperature.
8. Recover the cDNA by inverting the filter in a new tube and centrifuge at 7,300×*g* for 2 min.
9. Add appropriate volume of water to the filter column and centrifuge at 7,300×*g* for 1–2 min (see Note 8).
10. Repeat step 8.

3.1.6. Cap-Trapping with Streptavidin-Sepharose Beads

1. Mix 250 μL of streptavidin-sepharose beads and 2.5 μL of tRNA (20 μg/μL) and incubate at room temperature for 30 min or longer to coat the surface to avoid nonspecific binding.
2. Pour into MicroSpin column. Centrifuge at 13,000×*g* for 1 min and discard the flow-through.
3. Pour 250 μL of B&W buffer into the column and mix gently.
4. Centrifuge at 13,000×*g* for 1 min and discard the flow-through.
5. Pour 250 μL B&W buffer into the column and mix again by pipetting.

RNaseI treated solution	80 μL (adjust to 80 μL with water)
5 M NaCl	10 μL
0.5 M EDTA (pH 8.0)	10 μL
Washed beads	250 μL

6. Mix all of the above mentioned solutions with washed beads.
7. Keep at room temperature for 15 min or longer.
8. Centrifuge at 13,000×*g* for 10 s and discard the flow-through.

9. Go to washing steps.

   First wash: Pour 250 µL of B&W buffer into column and mix the solution.

   - Centrifuge at 13,000×*g* for 10 s and discard the flow-through.

   Second wash: Pour 250 µL of B&W buffer with 0.05% SDS into column and mix the solution.

   - Centrifuge at 13,000×*g* for 10 s and discard the flow-through.

   Third and fourth washes: Pour 250 µL of B&W buffer with 0.1% Tween 20 (0.1% final) into the column and mix the solution.

   - Centrifuge at 13,000×*g* for 10 s and discard the flow-through.

Fifth to tenth washes: Repeat steps in first wash.

3.1.7. Release cDNA from Beads

1. Add 100 µL of 50 mM NaOH to the column prepared in the previous step and incubate at 37°C for 10 min.
2. Centrifuge at 13,000×*g* for 10 s and collect the flow-through in a new tube.
3. Add 100 µL of 1 M Tris–HCl (pH 7.0) to the above flow-through for neutralization.
4. Repeat these collection steps four times. (Collect the flow-through in the same tube. And, it is not necessary to add 1 M Tris–HCl in repeat steps.)
5. Pour the solution into MicroconYM-100 to change the buffer and lower the volume.
6. Centrifuge at 600×*g* for 25 min.
7. Discard the flow-through and add 300 µL of water to the column again.
8. Centrifuge at 600×*g* for 25 min.
9. Recover the cDNA by inverting the filter in a new tube and centrifuge at 7,300×*g* for 2 min.
10. Add the appropriate volume of water to the filter column and centrifuge at 7,300×*g* for 1–2 min (see Note 8).
11. Repeat step 9.

3.2. Second Strand

3.2.1. Single-Strand Linker Ligation

1. Incubate 10 µL of cDNA at 65°C for 5 min.
2. Cool on ice for 2 min.
3. Prepare 5′ linker.

(a) Mix equimolar amount of the 5′ linker upper and lower oligonucleotides and dilute with water to a final concentration of 10 ng/μL.

(b) Add NaCl to yield 0.1 M final concentration.

(c) The annealing process can be performed on a thermal cycler at the following temperatures: 95°C for 5 min – 83°C for 5 min – 47°C for 5 min – 35°C for 5 min – 11°C hold (see Note 10).

5′ Linker (10 ng/μL)	1 μL
TAKARA DNA ligation Kit ver.2.1 II	10 μL
TAKARA DNA ligation Kit ver.2.1 I	20 μL

4. Add 5′ Linker (10 ng) and TAKARA DNA ligation Kit ver.2.1 II solution and mix well.
5. Add TAKARA DNA ligation Kit ver.2.1 I solution and mix gently (see Note 11).
6. Incubate at 16°C overnight.

3.2.2. GFX-CTAB Purification

SSLL solution	40 μL
Water	110 μL (see Note 12)
High SALT CTAB solution	375 μL

1. Mix all of the above mentioned solutions and incubate at 45°C for 20 min and keep at room temperature for 10 min. This is for binding cDNA fragments to CTAB.
2. Pour the solution into the GFX column. Centrifuge at 15,000 × *g* for 1 min and discard the flow-through.
3. Pour in 600 μL wash buffer. Centrifuge at 15,000 × *g* for 1 min and discard the flow-through.
4. Pour in 80% EtOH. Centrifuge at 15,000 × *g* for 1 min and discard the flow-through.
5. Place the column on a new 1.5 mL tube.
6. Add 55 μL of water and keep at room temperature for 1–2 min.
7. Centrifuge at 15,000 × *g* for 1 min.
8. Add 10 μL of water and centrifuge at 15,000 × *g* for 1 min.
9. Collect the eluent.

3.2.3. Second-Strand Synthesis

SSLL solution	X μL	(Adjust to under 62 μL) (see Note 13)
LA Taq buffer (10×)	10 μL	
25 mM $MgCl_2$	10 μL	2.5 mM
2.5 mM dNTPs	16 μL	0.4 mM
5′ primer (100 ng/μL)	1 μL (see Note 14)	100 ng
LA Taq (5 U/μL)	1 μL	
Water	Total volume = 100 μL	

1. Mix all of the above mentioned reagents on ice.
2. Elongation is performed on a thermal cycler at the following temperatures: 94°C for 3 min – 42°C for 5 min – 68°C for 20 min – 62°C for 2 min – 4°C hold.

3.2.4. ProK Treatment

1. Add 2 μL of ProK (20 mg/mL) and 4 μL 0.5 M EDTA for each 100 μL reaction solution.
2. Incubate at 45°C for 20 min and then cool on ice for 2 min.

3.2.5. GFX-CTAB Purification

1. Pour 200 μL of water and 1 μL of tRNA (20 μg) into the column and keep at room temperature for 1 min (see Note 15).
2. Centrifuge at 15,000 × *g* for 1 min and discard the flow-through.

dsDNA solution	106 μL
High SALT CTAB solution	265 μL

3. Mix all of the above mentioned solutions and incubate at 65°C for 10 min, then place at room temperature for 10 min.
4. Pour the solution into the GFX column. Centrifuge at 15,000 × *g* for 1 min and discard the flow-through.
5. Pour in 600 μL wash buffer. Centrifuge at 15,000 × *g* for 1 min and discard the flow-through.
6. Repeat step 5.
7. Pour in 80% EtOH. Centrifuge at 15,000 × *g* for 1 min and discard the flow-through.
8. Place a the column on a new 1.5 mL tube.
9. Add 35 μL of water and keep at room temperature for 1–2 min.
10. Centrifuge at 15,000 × *g* for 1 min (see Note 4).
11. Add 10 μL of water and centrifuge at 15,000 × *g* for 1 min.
12. Collect the eluate.

3.3. EcoP15I Digestion and linker ligation

3.3.1. EcoP15I Digestion

1. Check the concentration of cDNA (see Note 16).

dsDNA solution	X μL
NEbuffer3 (10×)	6 μL
BSA (100×)	0.6 μL
ATP (10×)	6 μL
10 mM Sinefungin	0.6 μL
EcoP15I (1 U/μL) (see Note 17)	1 μL
Water	Total volume = 60 μL

2. Mix all of the above mentioned reagents and incubate at 37°C for 3 h.
3. Add 0.6 μL of 1 M $MgCl_2$ (for final concentration of 10 mM).
4. Incubate at 65°C for 20 min, then cool on ice (see Note 18).

3.3.2. 3′ Linker Ligation

1. Prepare 3′ linker.
   (a) Mix equimolar amount of the upper and lower oligonucleotides and dilute with water to a final concentration of 100 ng/μL.
   (b) Add NaCl to yield 0.1 M final concentration.
   (c) The annealing process can be performed on a thermal cycler at the following temperatures: 95°C for 5 min – 83°C for 5 min – 47°C for 5 min – 35°C for 5min – 11°C hold (see Note 10).

EcoP15I digested solution	60.6 μL
Buffer (5×)	24 μL
3′ Linker (100 ng/μL; 3′ LinkerL)	1 μL
T4 DNA ligase (400 U/μL)	3 μL
Water	31.4 μL (see Note 19)

2. Mix all of the above mentioned reagents gently and incubate at 16°C overnight.

3.3.3. Purification from the 3′ Linker with Beads

1. Mix 10 μL of streptavidin-sepharose beads and 2.5 μL of tRNA (20 μg/μL) and incubate at room temperature for 30 min or longer to coat the surface to avoid nonspecific binding.
2. Pour into a MicroSpin column and centrifuge at 13,000 × *g* for 1 min, then discard the flow-through.
3. Pour 250 μL of B&W buffer into the column and mix gently.
4. Centrifuge at 13,000 × *g* for 1 min, then discard the flow-through.

5. Pour 50 μL B&W buffer into the column and mix again by pipetting.

Linker ligated solution	120 μL
5 M NaCl	14.5 μL
0.5 M EDTA (pH 8.0)	14.5 μL

6. Mix all of the above mentioned solutions with beads prepared in Subheading 3.3.3, steps 1–5.
7. Keep at room temperature for 1 h or longer.
8. Centrifuge at 13,000 × *g* for 10 s and discard the flow-through.
9. Go to washing steps.

   First wash: Pour 250 μL of B&W buffer into column and mix the solution.

   Centrifuge at 13,000 × *g* for 10 s and discard the flow-through.

   Second wash: Pour 250 μL of B&W buffer with 0.05% SDS into column and mix the solution.

   Centrifuge at 13,000 × *g* for 10 s and discard the flow-through.

   Third and fourth washes: Pour 250 μL of B&W buffer with 0.1% Tween 20 (0.1% final) into column and mix the solution.

   Centrifuge at 13,000 × *g* for 10 s and discard the flow-through.

   Fifth to tenth washes: Repeat steps in first wash.
10. After washing beads, pipette 50 μL of water into the column and collect beads in a new tube with a pipette (see Note 20).
11. Centrifuge at 13,000 × *g* for 1 min and discard the supernatant.
12. Add 20 μL of water (as template for PCR).

3.4. PCR Amplification and Sequencing

3.4.1. First PCR Check

Water	Up to 20 μL
HF buffer (5×)	4 μL
2.5 mM dNTPs	1.6 μL
100 μM Fprimer	0.2 μL
100 μM Rprimer	0.2 μL
Template	0.2–0.8 μL
Phusion (2 U/μL)	0.2 μL

1. Mix template and all of the above mentioned reagents on ice.
2. PCR is performed on a thermal cycler at the following temperatures: 98°C for 30 s – (98°C for 10 s – 60°C for 10 s) X x cycles – 4°C hold (see Note 21).
3. Take out a tube at each step you would like to check and run on 12% PAGE gel.

3.4.2. First PCR

Water	Up to 50 μL
HF buffer (5×)	10 μL
2.5 mM dNTP	4 μL
100 μM Fprimer	0.5 μL
100 μM Rprimer	0.5 μL
template	0.5–2 μL
Phusion (2 U/μL)	0.5 μL

1. Mix template and all of the above mentioned reagents on ice.
2. PCR is performed on a thermal cycler at the following temperatures: 98°C for 30 s – (98°C for 10 s – 60°C for 10 s) X cycles – 4°C hold (X is the number of PCR cycles. See note 22).

3.4.3. proK Treatment and EtOH Precipitation

1. Mix 400 μL of PCR-reacted solution, 8 μL of 0.5 M EDTA (pH 8.0), and 8 μL of proK.
2. Incubate at 45°C for 15 min.
3. Mix the DNA solution with 41.6 μL of 5 M NaCl, 2 μL of Glycogen (1 μg/μL), and 920 μL of cold 99.5% EtOH.
4. Vortex well and chill at –20°C (or –80°C) for 15 min.
5. Centrifuge at 15,000×*g* for 15 min and finish the centrifugation, then discard the supernatant.
6. Add 500 μL of cold 80% EtOH.
7. Centrifuge at 15,000×*g* for 10 min and finish the centrifugation, then discard the supernatant.
8. Repeat steps 6 and 7.
9. Dry the pellet.
10. Dissolve the pellet with 20 μL of water.
11. Add loading dye.
12. Apply 10 μL of sample per lane to a 5% agarose gel.
13. Run the electrophoresis.
14. Cut out the desired fragment and collect in a 1.5-mL tube.

3.4.4. Purification and Sequencing

1. Measure the gel slice.
2. Add β-agaraseI (1×) buffer for 1% agarose final.
3. Add β-agaraseI (1 U/μL), 1 U/200 μL solution.
4. Incubate at 37°C for 1 h.
5. Place on ice and check whether the gel has been completely digested.
6. Add an equal volume of PCI and mix thoroughly.
7. Centrifuge at 15,000×*g* for 10 min.
8. Transfer the upper layer to a new tube (see Note 23).

9. Repeat steps 7 and 8.
10. Add an equal volume of CIA and mix thoroughly, then centrifuge at 13,000 ×*g* for 10 min.
11. Transfer the upper layer (DNA solution) to a new tube.
12. Mix the DNA solution with 41.6 µL of 5 M NaCl, 2 µL of Glycogen (1 µg/µL), and 920 µL of cold 99.5% EtOH.
13. Vortex well and chill at –20°C (or –80°C) for 15 min.
14. Centrifuge at 15,000 ×*g* for 15 min and finish the centrifugation, then discard the supernatant.
15. Add 500 µL of cold 80% EtOH.
16. Centrifuge at 15,000 ×*g* for 10 min, then discard the supernatant.
17. Repeat steps 15 and 16.
18. Dry the pellet.
19. Dissolve the pellet with 20 µL of water.
20. Run on the Genome Analyzer.
21. The sequencing result is basecalled, mapped, and visualized in silico (Fig. 2).

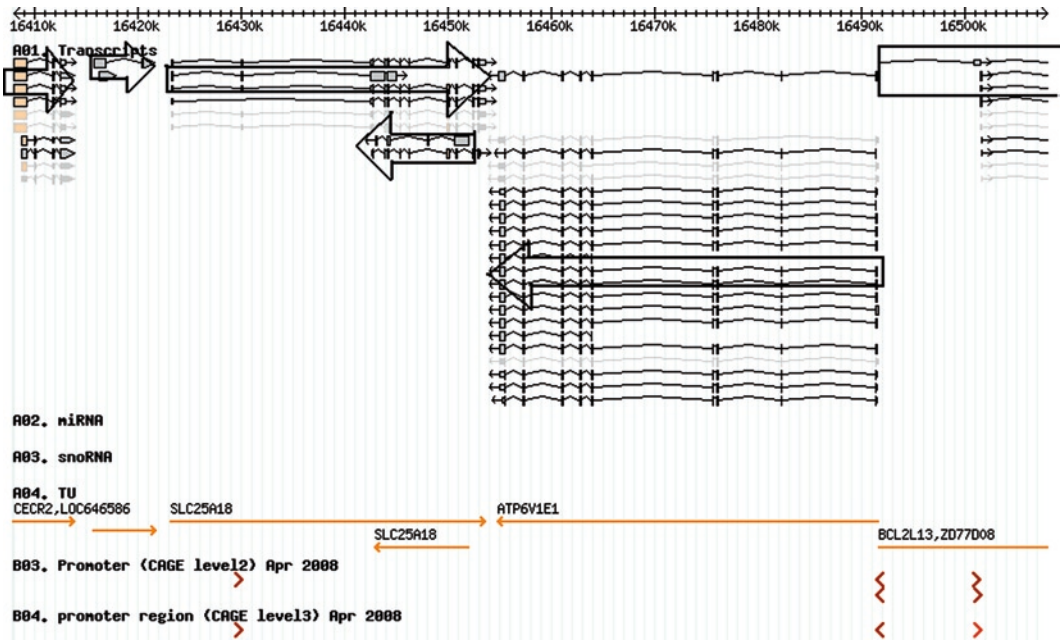


Fig. 2. An example of the result: the sequenced Redundant CAGE Tags are mapped on the genome in silico. This result indicates that there are at least four distinct transcription units between 16,410 and 165,000 kb on this chromosome. Three of five are from plus, and two are from minus strands in this region. It also suggests that the minus strand between 16,450 and 16,490 kb exhibits an especially high expression level. This genetic mapping is drawn by the Genomic Viewer at the Web page of GSC RIKEN Yokohama Institute (http://fantom.gsc.riken.go.jp/).

4. Notes

1. Do not use a metal spoon.
2. Use of PAGE purified linker is strongly recommended. Linker with a lower grade of purification may contain incomplete biotin-added linker, which causes loss of sample after 3′ linker ligation, or a few base pairs shorter linker, which causes linker dimer formation.
3. The ratio of RT primer should be "Random to oligo dT = 4:1."
4. If the concentration of RNA is low so that it exceeds the total volume of the enzyme mix, you should adjust the quantity of the buffer, dNTPs, or sorbitol/trehalose.
5. When a white precipitate from CTAB is observed, it is best to transfer the supernatant to a new tube.
6. $NaIO_4$ should be prepared fresh prior to each use to avoid light-exposure damage.
7. It is necessary to check how the fragment has flowed and adjust the volume of water for the next try.
8. Measure the volume of recovered solution before adding appropriate volume of water to decide the volume of water in this and the next step because the volume of recovered solution is decided in the next section.
9. Adjust Biotin (long arm) Hydrazide immediately prior to use since Biotin Hydrazide is highly susceptible to moisture.
10. Set the reaction at −0.1°C/s. This reaction requires slow temperature drops.
11. Mix gently to avoid denaturing an enzyme contained in TAKARA DNA ligation Kit ver.2.1 I.
12. The function of water in this step is to adjust the stickiness of PEG contained in the TAKARA DNA ligation Kit.
13. SSLL solution should be adjusted to be below 62 μL because the total volume is 100 μL.
14. In order to avoid misannealing the N6 of the 5' linker to the cDNAs, a 5' primer should be added.
15. Make it easy to flow short fragments by binding tRNA to glass filter beforehand.
16. EcoP15I can digest 100 ng of dsDNA per 1 U. Check the concentration of cDNA and adjust the volume of dsDNA solution in the next step.
17. The initial concentration was 10 U/μL, so it should be diluted with 1× buffer before using.

18. It can heat-inactivate the enzyme.
19. The total volume should be higher than double the volume of EcoP15I reaction solution.
20. The volume can exceed 50 μL, but beads have to be completely collected.
21. Overcycled PCR causes bias by interfering with amplification, by forming 3′ linker and 5′ linker dimers. Optimization of the adequate number of PCR cycles will be required for each sample. We usually test with four different PCR cycles with negative control. Testing with 10, 12, 14, and 16 cycles are recommended. On the optimal cycle, you can detect a faint single band by agarose gel electrophoresis.
22. Adopt the PCR cycle optimized in the previous section. In order to collect enough PCR products, repeat the PCR reaction eight times.
23. Avoid taking the upper layer that contains proteins and contaminants.

Acknowledgment

This research is supported by Research Grant for RIKEN Omics Science Center from MEXT to YH.

References

1. Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., Fukuda, S., Sasaki, D., Podhajska, A., Harbers, M., Kawai, J., Carninci, P., and Hayashizaki, Y. (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A* **100**, 15776–81.
2. Maeda, N., Kasukawa, T., Oyama, R., et al. (2006) Transcript annotation in FANTOM3: mouse gene catalog based on physical cDNAs. *PLoS Genet* **2**, e62.
3. Hayashizaki, Y. and Carninci, P. (2006) Genome Network and FANTOM3: assessing the complexity of the transcriptome. *PLoS Genet* **2**, e63.
4. de Hoon, M. and Hayashizaki, Y. (2008) Deep cap analysis gene expression (CAGE): genome-wide identification of promoters, quantification of their expression, and network inference. *Biotechniques* **44**, 627–32.

Chapter 11

Linking Emulsion PCR Haplotype Analysis

James G. Wetmur and Jia Chen

Abstract

The experimental measurement of haplotype requires the determination of two or more genotypes on the same DNA molecule. Because such measurements are much more complicated than measurements of genotypes, haplotypes are typically inferred using population data for linkage disequilibrium between the markers of interest. We have developed a method for molecular haplotyping, linking emulsion PCR (LE-PCR), and have demonstrated that the method is sufficiently robust to determine haplotypes for multiple markers in a population setting. LE-PCR uses emulsion PCR to isolate single template molecules for simultaneous PCR of widely spaced markers and uses linking PCR to fuse these amplicons into one short amplicon, which maintains the phase of the markers. LE-PCR is illustrated for polymorphisms in human paraoxonase 1 (*PON1*) that have been shown to affect transcriptional activity and substrate specificity in the detoxification of organophosphates.

Key words: Linking PCR, Emulsion PCR, Genotype, Phenotype, Haplotype, Paraoxonase-1

1. Introduction

Linking emulsion PCR (LE-PCR) merges emulsion PCR (ePCR) with linking (fusion) PCR. ePCR is a single-molecule technique where one template molecule may be amplified within a single aqueous droplet in an oil–water emulsion. PCR in emulsions was developed for directed molecular evolution and has found many and varied applications in the field, with selection criteria including catalysis, binding, and regulation. ePCR also has many applications beyond directed evolution, including cell-free cloning, where multiple amplicons within a droplet, often attached to a microbead, may be made available for genotyping or high-throughput sequencing. The applications to DNA sequencing have been aided by developments in digital microfluidics, where it is possible to create and manipulate monodisperse emulsions. All of these methods have been reviewed extensively (1–4).

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_11, © Springer Science+Business Media, LLC 2011

Linking (also called fusion) PCR is a technique for merging two amplicons into one amplicon (5, 6). We developed LE-PCR by combining linking PCR with ePCR to enable preservation of phase during linkage of two widely separated polymorphic markers from the same DNA duplex into a minichromosome (7). This chapter is an update of a chapter completed in 2005 for this same series (8).

LE-PCR is one of several approaches to molecular haplotyping. If haploid samples were available, as with sperm (9), somatic cell hybrids (10) or cloned DNA cosmids, YACs, etc. (11), then the haplotype could be determined by genotyping. Another approach would be to obtain a haploid sample by hybrid selection (12). The haplotype can be determined over the entire length of the haploid sample. Sperm typing is limited to males, and the other methods are generally too complex and expensive to apply to a population-based association study. Long-range allele-specific PCR is difficult because of the requirements of both long range using an editing DNA polymerase and allele-specific amplification and has been limited to 10 or at the outside 20 kb (13, 14). With LE-PCR, large marker separations can be accommodated, as DNA fragments of 100–150 kb have been shown to be resistant to shear during emulsion preparation (15). Polony haplotyping, which replaces the emulsion with a gel to isolate template molecules, also works just like LE-PCR for haplotyping large molecules but requires dedicated instrumentation that may be unavailable to most laboratories (16). Finally, again starting with isolated genomic DNA, the oldest single-molecule method is limiting dilution, the basis of digital PCR (17–19). Limiting dilution PCR, like sperm typing, requires extreme precautions to prevent contamination. In addition, one template molecule must support more than one PCR in a large volume. Although limiting dilution PCR is a straightforward and competitive technology, LE-PCR may be more robust and less error-prone for population-based research.

2. Materials

2.1. Human Genomic DNA

Any source of human DNA may be used. However, because polymorphic markers for haplotyping must be on the same molecule, mechanical shear should be minimized. Salt concentrations of 0.4 M or higher will minimize shear if the samples must be frozen. Repeated cycles of freezing and thawing should be avoided. Inclusion of 1 mM EDTA, pH ≥ 7.8, in storage buffers will inhibit DNases, although small numbers of single-strand breaks will not affect LE-PCR as the single molecules are isolated before denaturation. In the example in this chapter, DNA was obtained from

a study population of the Mount Sinai Children's Environmental Health Center, a study to assess prospectively infant growth and neurodevelopment associated with pesticide exposure in urban New York City. The study protocol was approved by the Mount Sinai School of Medicine Institutional Review Board. The study population consisted of pregnant women of multiethnic origin (Caucasian, African-American, and Hispanic of Caribbean origin) at 26–30 weeks of gestational age. Leukocyte DNA was isolated from blood as previously described (20).

2.2. Oligonucleotides as Illustrated for PON1 -909g>c and Q192R

Of course, oligonucleotides for haplotyping sets of markers by LE-PCR must be designed for the specific markers of interest. The primer sequences below are given for illustration purposes only.

1. All primers were synthesized by IDT, Inc.
2. External primers for amplification across *PON1 -909g>c* and *Q192R*:

   CAAAATCAAATCCTTCTGCCACCACTCGAA and ACATGGAGCAAATCATTCACAGTAA, respectively.
3. Linking primers for *PON1 -909g>c* and *Q192R* (5′-biotinylated):

   Bio-AAAGTGCTCAGGTCCCACACTGATAATGGGGCATTTGAGTAA and Bio-GCCCCATTATCAGTGTGGGACCTGAGCACTTTTATGGCACAA, respectively.
4. Capping oligonucleotides for *PON1 -909g>c* and *Q192R* (3′-phosphorylated):

   AAAAAAGCCCCATTATCAGTG-P and AAAAAAAAAGTGCTCAGGTCCCA-P.
5. qASPCR primers:

   192T: CAAATACATCTCCCAGGATT and 192C: CAAATACATCTCCCAGGATC

   -909C: GCAGACAGCAGAGAAGAGAC and -909G: GCAGACAGCAGAGAAGAGAG.

2.3. Buffers (All 1×)

1. Taq: 10 mM Tris–HCl, pH 8.0, 50 mM KCl.
2. NX: 100 mM NaCl, 1% Triton X-100, 10 mM Tris–HCl, pH 7.5, 1 mM EDTA.
3. B&W: 10 mM Tris–HCl, pH 7.5, 1 mM EDTA, 2 M NaCl.
4. qASPCR: 1× Taq buffer, 1.5 mM $MgCl_2$, 200 μM each dNTP, 2% glycerol, 1× BSA (NEB), and 1× SYBR green (Invitrogen – Molecular Probes).

2.4. Emulsion Constituents

The emulsion constituents listed below were used for our published study (7). The reader is directed to alternative formulations for Finnzymes (15) and for PfuTurbo polymerase (21).

1. Oil phase final concentrations: 4.5% Span 80 (#85548, Fluka), 0.4% Tween 80 (#S-8074, Sigma), 0.05% Triton X-100 (#T-9284, Sigma) made up to 100% with mineral oil (#M-3516, Sigma).
2. Aqueous phase final concentrations: 1× Taq buffer, 300 μM each dNTP, 2.5 mM $MgCl_2$, 50 μM Me_4NCl, 1 μM each external primer, 0.1 μM each linking primer, 100 mU/μL Amplitaq Gold (Applied Biosystems), and 1 ng/μL human genomic DNA.

2.5. Capping Reaction Constituents (Final Concentrations)

1. 1× Taq buffer, 1.5 mM $MgCl_2$, 200 μM each dNTP.
2. 1 μM each capping oligonucleotide.
3. 5 U/40 μL Taq DNA polymerase (Promega, not hot start).

2.6. qASPCR Constituents (Final Concentrations)

1. 1× qASPCR buffer.
2. 1 μM each qASPCR primer.
3. 2.5 U/20 μL Amplitaq Gold DNA polymerase (Applied Biosystems).

2.7. Other Materials

1. PCR purification kit (Qiagen).
2. Dynabeads Myone Streptavidin (Dynal Biotech).

3. Methods

The overall logic for LE-PCR is illustrated in Fig. 1. Two amplicons are produced spanning the linked polymorphisms (*A* and *B*) within an aqueous droplet in an emulsion. Molecular haplotypes need to be determined for two polymorphic loci only if they both are heterozygous (*A/a* and *B/b*) or at most 25% of all samples. Linking PCR connects these amplicons into minichromosomes, preserving the phase information of the two polymorphisms. For example, in a total population of 378, 77 subjects had compound heterozygous genotypes at *PON1 -909g>c* and *Q192R*, thus ambiguous haplotypes (7). The minichromosomes will represent both haplotypes of either diplotype I or II. Now, consider a DNA region with three polymorphisms of interest where any two or all three loci may be heterozygous, again as illustrated for *PON1*. The *-909g>c* polymorphism is in nearly complete LD with a functional promoter polymorphism at -108 that affects the level of transcription; *Q192R* affects substrate specificity and *L55M* may alter the lifetime of the enzyme. To include *L55M* in the analysis required establishment of LE-PCRs for each pair of loci, in this case 3, with some increase in the fraction of the samples requiring molecular haplotyping. In the rarer samples where all three loci are heterozygous, two measurements are required but the third

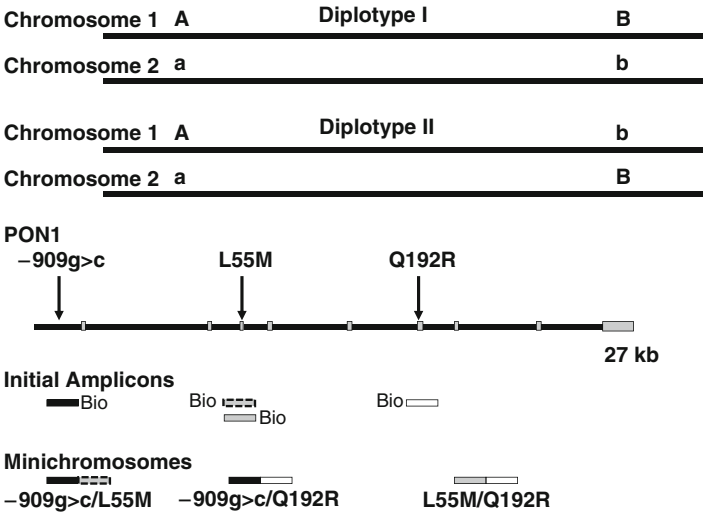


Fig. 1. Linking emulsion PCR design. *Top*: Diplotypes with two widely spaced heterozygous markers (*A*/*a* and *B*/*b*). Diplotype I formed from haplotypes *AB* and *ab*. Diplotype II formed from haplotypes *Ab* and *aB*. *Bottom*: Illustration with human *PON1* drawn to scale. Three polymorphic markers are indicated, *-909g>c* in the 3′ untranslated region, *L55M* in exon 3 and *Q192R* in exon 6. Initial amplicons are shown for the three polymorphic sites, as are the three possible minichromosomes fusing initial amplicons.

to be inferred. The initial amplicons and the structure of the resultant three *PON1* minichromosomes are shown. By comparing the predictive power of molecular vs. inferred haplotypes to the phenotype, we have demonstrated the utility of molecular haplotyping by LE-PCR in population studies by demonstrating haplotype-specific enzymatic activity (7). In this chapter, we limit our illustrations to the two loci *PON1 -909g>c* and *Q192R* and do not consider measurements of phenotypes.

3.1. Primer Design

1. Emulsion PCR requires two primers for each of the two amplicons as illustrated at the top of Fig. 2. These initial amplicons should be limited to 200 nucleotides.
2. The external primers are typical PCR primers of about 25 nt. It is our experience that primers containing approximately 50% GC and ending in AA are optimal.
3. The internal primers are the linking primers. The design of overlapping partially complementary linking primers is shown together with the example for *PON1 -909g>c* and *Q192R*. Thirty-two of 42 nucleotides are complementary beginning at the 5′ end. The 26 nucleotides at the 3′ end are complementary to the template and act as the primer. The 16 nucleotides at the 5′ end of each linking primer are derived by complementarity to the other linking primer and hence the template for the other amplicon. They must carry a 5′-biotin to allow subsequent separation of unlinked amplicons.

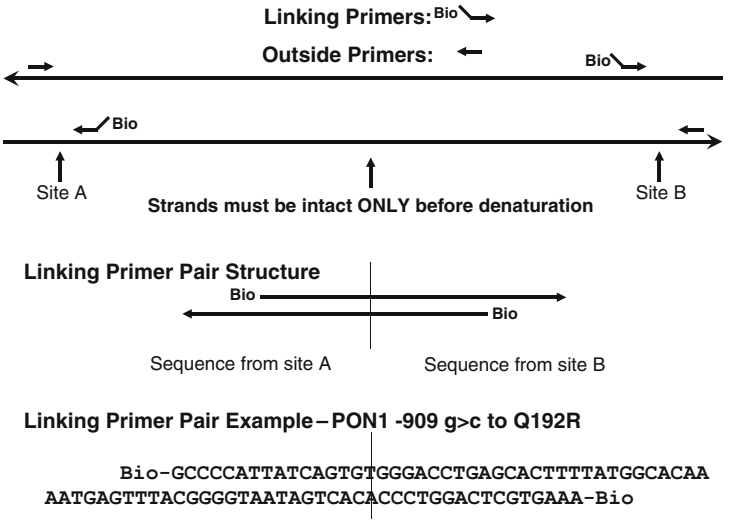


Fig. 2. Linking emulsion PCR primers. *Top*: External and biotinylated linking primers on denatured template strands for synthesis of initial amplicons in an emulsion droplet. *Center*: Design of the linking primers. After two PCR cycles, the 3′ end of the complementary strand of each amplicon may act as a primer on the other amplicon, permitting synthesis of a minichromosome. The external primers are added in excess leading to depletion of the linking primers and bias toward minichromosome formation. *Bottom*: Illustration of linking primer complementarity for human *PON1 -909g>c* and *Q192R*.

3.2. Emulsion Formation

1. The template between the amplicons must be intact in the aqueous droplets before PCR. We have shown that emulsification had no effect on template integrity over the 15 kb between *PON1 -909g>c* and *Q192R*. We measured the water droplets to be in the order of 10 μm and less in diameter. By limiting template concentration to 1 ng/μL (~300 haploid genome equivalents/μL), the probability of having more than one template per droplet is less than 1%.
2. Both the oil phase and aqueous phase are assembled from the constituents listed in Subheading 2.4, items 1 and 2, respectively. Examples of external and linking primer sequences are given for *PON1 -909g>c* and *Q192R* in Subheading 2.3, items 2 and 3, respectively. The human template DNAs to be haplotyped are heterozygous for both linked polymorphisms.
3. Emulsification involves vortexing one part aqueous phase and two parts oil phase (typical volume 150 μL) for 5 min. We have found that tubes in a foam holder strapped to a EW-04725-10 Thermo Scientific Vortex Shaker for 5 min will form stable emulsions. See Note 1. Emulsions may also be produced by stirring (21).
4. Transfer the tubes to a PCR machine and overlay with mineral oil.

3.3. LE-PCR

1. For the example used in this chapter, the PCR cycling conditions were 30 cycles for 1 min at 67°C, 1 min at 60°C, 30 s at 94°C following incubation at 95°C for 9 min to activate the polymerase and followed by a final incubation for 7 min at 60°C.
2. The external primers need to be present in tenfold excess over the linking primers. The PCR conditions were chosen so that at later cycles, 1 min at 67°C favored extension from longer primers, in this case first from the dilute linking primers and later from the amplicon strands themselves acting as primers, as required for linking.

3.4. PCR Cleanup

1. Add three volumes of NX buffer to five volumes of emulsion. Vortex for 20 s. Separate the phases in a microcentrifuge and remove most of the oil.
2. Transfer the aqueous phase to a Qiagen PCR purification kit and use according to the manufacturer's instructions. The Qiagen PCR purification kit tolerates oil carryover. Elute in 40 μL.

3.5. Removal of Biotinylated Primers and Unlinked Amplicons

1. The products of LE-PCR are illustrated at the top of Fig. 3, including the desired minichromosomes, contaminating unlinked initial amplicons spanning the two heterozygous polymorphic sites and contaminating single-stranded external primer runoff products from the two amplicons, a consequence

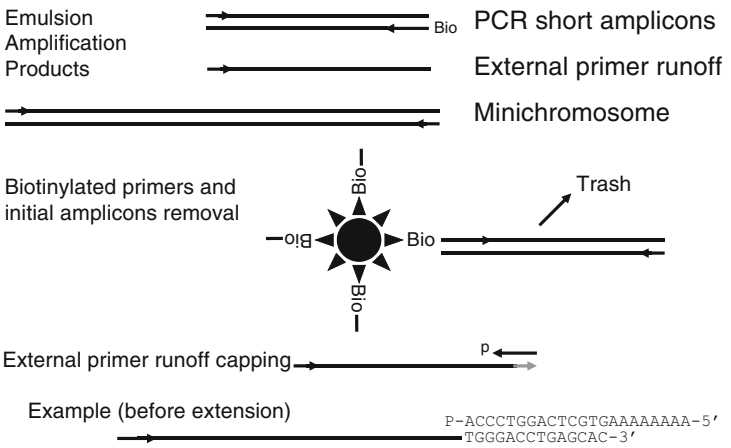


Fig. 3. Postemulsion PCR steps. *Top*: LE-PCR products include the desired minichromosomes, unlinked initial amplicons and single-stranded external primer runoff products formed due to excess external vs. linking primers. *Middle*: Magnetic bead-based removal of biotin-containing initial amplicons and any unreacted linking primers. *Bottom*: Capping of single-stranded external primer runoff products. Incubation with a nonediting DNA polymerase adds the gray sequence at the 3′ end. In the example, eight T residues would be added. The 3′-phosphate on capping oligonucleotides prevents their action as primers.

of the excess of external primers over linking primers. Removal of one of the two contaminants, the unlinked initial amplicons, is facilitated by the presence of the biotins on these amplicons and their absence on the minichromosomes. Because of the length and self-complementarity of the linking primers, unreacted linking primers may also remain after the Qiagen PCR purification step, but they also contain biotins. Removal of all biotinylated nucleic acids is illustrated in the center of Fig. 3.

2. Wash 3 µL Dynabeads Myone streptavidin (Dynal Biotech) three times in B&W buffer and once in Taq buffer. Resuspend the beads in the 40 µL eluate from the Qiagen PCR cleanup and add 4 µL 10× Taq buffer. Incubate at room temperature for 30 min, magnetize and retain the supernatant.

3.6. Capping of Runoff Products

1. After breaking the emulsion, the single-stranded external primer runoff products from the two amplicons remain in solution with the desired minichromosome products. In any subsequent PCR, these contaminants could act as primers for minichromosome formation where phase would not be preserved as it had been in LE-PCR with isolated template molecules in emulsion droplets. Instead of removing these contaminants, we chose to disable their ability to act as primers.

   The bottom of Fig. 3 illustrates the use of capping oligonucleotides. The capping oligonucleotides are complementary to the 3′ end of the runoff products and contain an extended 5′ sequence. The extension of the 3′ end of the runoff products prevents them from acting as primers. The 3′-phosphate synthesized onto the capping oligonucleotides prevents their action as primers when using a nonediting DNA polymerase such as Taq. After the postemulsion purification steps in Subheadings 3.5 and 3.6, PCR may be carried out on minichromosome templates without loss of phase information.

2. Capping reactions are assembled as described in Subheading 2.5. The oligonucleotides in Subheading 2.2, item 4 were used for *PON1 -909g>c* and *Q192R* runoff products as illustrated for *Q192R* in the figure. Use "vanilla" Taq and not a hot-start DNA polymerase. Incubate at 55°C for 30 min. This step completes the postemulsion steps for LE-PCR.

3.7. Determination of Haplotypes by qASPCR

1. As noted at the top of Fig. 1, minichromosome diplotypes may be of type I or II. An allele-specific PCR primer set for diplotype I could be specific for *A* and *B*, one of the haplotypes,

or for *a* and *b*, the other. Similarly, an allele-specific PCR primer set for diplotype II could be specific for *A* and *b* or for *a* and *B*. All four qASPCRs should be performed on each sample. Each qASPCR uses 2 μL of capped product in a 20 μL PCR and is assembled as noted in Subheading 2.6. Ideally, the assays are carried out at least in duplicate. For illustration, the primers for *PON1 -909g>c* and *Q192R* are given in Subheading 2.2, item 5.

2. In our example, PCR was carried out in a LightCycler (Roche) with cycling 1 min at 55°C, 1 min at 72°C, 30 s at 94°C after an initial incubation at 95°C for 9 min to activate the polymerase. Ct values were determined by the LightCycler using the second derivative algorithm to analyze results for of the four primer pairs.
3. Any real-time PCR instrument should be equally effective.

3.8. Data Analysis

1. All LE-PCRs will lead to the formation of one minichromosome diplotype of two haplotypes preserving the phase information of the polymorphic alleles on the two template chromosomes.
2. The calculations assume all qASPCR primers are equally efficient when amplifying PCR matched templates, as was the case for our four examples. If not, appropriate corrections must be employed.
3. Calculate the average Ct for the first possible haplotype pair: in the example, *192T/-909c* plus *192C/-909g*.
4. Calculate the average Ct for the second possible haplotype pair: in the example, *192T/-909g* plus *192C/-909c*.
5. Subtract to obtain ΔCt.
6. Require that BOTH Ct values for the first possible haplotype pair (e.g., both amplifications with allele-specific primers for *192T/-909c* and with allele-specific primers for *192C/-909g*) be less than BOTH Ct values for the second possible haplotype pair (e.g., *192T/-909g* and *192C/-909c*) *or* BOTH Ct values for the first possible haplotype pair be greater than BOTH Ct values for the second possible haplotype pair. Thus, the four qASPCR measurements consistently favor one haplotype pair (diplotype).
7. Require that $\Delta Ct > 1$ or $\Delta Ct < -1$. Thus, the extent to which the qASPCR measurements favor one haplotype pair is well above the experimental error of the technique.
8. If both conditions 6 and 7 are met, call the haplotype based on the lowest Ct values.
9. If both conditions 6 and 7 are not met, see Note 2.

4. Notes

1. It is important that the tubes be in contact with the platform. Devices hanging the tubes around the sides produce poor emulsions.
2. LE-PCR failures are characterized by observing deltaCt values close to 0. Early in the development of this method, LE-PCR failures occurred about 20–30% of the time, but after practice in forming the emulsions, the failure rate dropped to less than 5%. We found that all such failures occurred in the emulsion PCR step and could not be rescued by repeating the purification and assay steps. If a haplotype cannot be called, repeat the entire LE-PCR experiment.

Acknowledgments

This work was supported by grants R21 ES011634 and P01 ES09584 from the National Institute of Environmental Health Sciences and RD831-711 from the Environmental Protection Administration.

References

1. Griffiths, A.D. and Tawfik, D.S. (2006) Miniaturising the laboratory in emulsion droplets. *Trends Biotechnol* **24**, 395–402.
2. Kelly, B.T., Baret, J.C., Taly, V., and Griffiths, A.D. (2007) Miniaturizing chemistry and biology in microdroplets, *Chem Commun* (*Camb*) **18**, 1773–88.
3. Taly, V., Kelly, B.T., and Griffiths, A.D. (2007) Droplets as microreactors for high-throughput biology. *Chembiochem* **8**, 263–72.
4. Zhang, Y. and Ozdemir, P. (2009) Microfluidic DNA amplification – a review. *Anal Chim Acta* **638**, 115–25.
5. Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K., and Pease, L.R. (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**, 61–8.
6. Yon, J. and Fried, M. (1989) Precise gene fusion by PCR. *Nucleic Acids Res* **17**, 4895.
7. Wetmur, J.G., Kumar, M., Zhang, L., Palomeque, C., Wallenstein, S., and Chen, J. (2005) Molecular haplotyping by linking emulsion PCR: analysis of paraoxonase 1 haplotypes and phenotypes. *Nucleic Acids Res* **33**, 2615–19.
8. Wetmur, J.G. and Chen, J. (2008) An emulsion polymerase chain reaction-based method for molecular haplotyping. *Methods Mol Biol* **410**, 351–61.
9. Li, H.H., Gyllensten, U.B., Cui, X.F., Saiki, R.K., Erlich, H.A., and Arnheim, N. (1988) Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* **335**, 414–7.
10. Douglas, J.A., Boehnke, M., Gillanders, E., Trent, J.M., and Gruber, S.B. (2001) Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. *Nat Genet* **28**, 361–4.
11. Burgtorf, C., Kepper, P., Hoehe, M., Schmitt, C., Reinhardt, R., Lehrach, H., and Sauer, S. (2003) Clone-based systematic haplotyping (CSH): a procedure for physical haplotyping of whole genomes. *Genome Res* **13**, 2717–24.
12. Guo, Z., Hood, L., Malkki, M., and Petersdorf, E.W. (2006) Long-range multilo-

cus haplotype phasing of the MHC. *Proc Natl Acad Sci U S A* **103**, 6964–9.

13. Michalatos-Beloin, S., Tishkoff, S.A., Bentley, K.L., Kidd, K.K., and Ruano, G. (1996) Molecular haplotyping of genetic markers 10 kb apart by allele-specific long-range PCR. *Nucleic Acids Res* **24**, 4841–3.
14. McDonald, O.G., Krynetski, E.Y., and Evans, W.E. (2002) Molecular haplotyping of genomic DNA for multiple single-nucleotide polymorphisms located kilobases apart using long-range polymerase chain reaction and intramolecular ligation. *Pharmacogenetics* **12**, 93–9.
15. Turner, D.J., Shendure, J., Porreca, G., Church, G., Green, P., Tyler-Smith, C., and Hurles, M.E. (2006) Assaying chromosomal inversions by single-molecule haplotyping. *Nat Methods* **3**, 439–45.
16. Mitra, R.D., Butty, V.L., Shendure, J., Williams, B.R., Housman, D.E., and Church, G.M. (2003) Digital genotyping and haplotyping with polymerase colonies. *Proc Natl Acad Sci U S A* **100**, 5926–31.
17. Ruano, G., Kidd, K.K., and Stephens, J.C. (1990) Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules. *Proc Natl Acad Sci U S A* **87**, 6296–300.
18. Stephens, J.C., Rogers, J., and Ruano, G. (1990) Theoretical underpinning of the single-molecule-dilution (SMD) method of direct haplotype resolution. *Am J Hum Genet* **46**, 1149–55.
19. Vogelstein, B. and Kinzler, K.W. (1999) Digital PCR. *Proc Natl Acad Sci U S A* **96**, 9236–41.
20. Chen, J., Kumar, M., Chan, W., Berkowitz, G., and Wetmur, J.G. (2003) Increased influence of genetic variation on PON1 activity in neonates. *Environ Health Perspect* **111**, 1403–9.
21. Williams, R., Peisajovich, S.G., Miller, O.J., Magdassi, S., Tawfik, D.S., and Griffiths, A.D. (2006) Amplification of complex gene libraries by emulsion PCR. *Nat Methods* **3**, 545–50.

Chapter 12

PAP-LMPCR: An Improved, Sequence-Selective Method for the In Vivo Analysis of Transcription Factor Occupancy and Chromatin Fine Structure

Richard Ingram, Arthur Riggs, and Constanze Bonifer

Abstract

In vivo footprinting and ligation-mediated PCR (LMPCR) are well-established methods for the examination of the chromatin structure of eukaryotic genes. Here, we describe an improved method (pyrophosphorolysis activated polymerization LMPCR or PAP-LMPCR) that overcomes the shortfalls of previous methods by being capable of reading through sequences that up to now were refractory to this type of analysis. This includes dinucleotide repeat sequences or GC-rich regions. We also describe conditions capable of distinguishing between different alleles, thus enabling the simultaneous analysis of monoallelically expressed genes without having to employ interspecies hybrids.

Key words: In vivo footprinting, LM-PCR, Chromatin structure analysis, Allele-specific chromatin structure, X-inactivation, CSF1R

1. Introduction

The activation of any gene is associated with alterations in chromatin structure. These may be large scale, such as the modification of extended regions of chromatin through changes in histone acetylation and methylation, or more localised through the binding of individual transcription factors. These factors, in turn, recruit chromatin remodelling activities capable of altering nucleosome occupancy and composition (1, 2). These changes can be measured by treating chromatin with DNA modifying agents such as dimethyl sulphate (DMS) (see Note 1) or nucleases such as dexoyribonuclease I (DNase1) or micrococcal nuclease (MNase). Treatment of living or permeabilised cells with one of these agents, followed by the use of LM-PCR, allows the researcher to visualise

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_12, © Springer Science+Business Media, LLC 2011

the changes in the fine structure of chromatin that take place during the up- or downregulation of a gene (3). Our methods of conventional LM-PCR have been published previously (4–6), and a very detailed description including troubleshooting can be found in Tagoh et al. (7). This method is capable of working on a large proportion of the genome. However, as with all PCR-based methods, there are some sequences for which it is difficult to design good primers because they are either of low sequence complexity or have a high GC content. In addition, in some cases, such as for imprinted genes and genes on the X chromosome, the two alleles within the nucleus adopt different chromatin structures, and it would be desirable to be able to examine just one allele at a time. To solve this problem, we utilised pyrophosphorolysis to increase the specificity of LM-PCR when dealing with "difficult" sequences. Pyrophosphorolysis-activated polymerization (PAP) (8–10) is a PCR-type amplification that utilises 3′-blocked primers that can be activated by pyrophosphorolysis while annealed to the complementary DNA strand in the presence of pyrophosphate. During DNA polymerisation, the incorporation of NTPs into the growing chain releases pyrophosphate. In the presence of high concentrations of pyrophosphate, this polymerisation is a reversible reaction, thus removing rather than adding nucleotides. In the presence of pyrophosphate some DNA polymerases can remove a blocking nucleotide, such as acycloNMP or ddNMP from the 3′ end of a primer (11, 12). The use of a blocked primer increases specificity, as removal of the blocked nucleotide by pyrophosphorolysis only occurs if the primer is perfectly annealed; any mismatches at or near the 3′ end of the primer prevents pyrophosphorolysis from occurring, and hence, elongation does not take place (12). Hence, this methodology, whilst reducing sensitivity somewhat (see Note 2), selects for the correct LM-PCR product, since other products cannot be amplified. The principle of this reaction is outlined in Fig. 1.

1.1. Overview of the PAP-LMPCR Procedure

A detailed description of how to treat cells with DMS, DNAseI, and MNase can be found in Tagoh et al. (7). Once DNA has been purified from treated cells, it is used in a primer extension reaction with a terminated primer. Although there are a number of possible nucleotide modifications that can be used to block elongation, dideoxyterminated primers are currently commercially available and hence are used here. Should the user wish to produce their own blocked oligos, the method can be found in (10) and (11). In conventional LM-PCR, the primer extension step uses Vent exonuclease however, this enzyme does not efficiently remove the blocking nucleotide. In PAP-LMPCR, the best results were obtained with Therminator™ polymerase, which is better at handling modified nucleotides (13, 14) (see Note 3), Primer extension

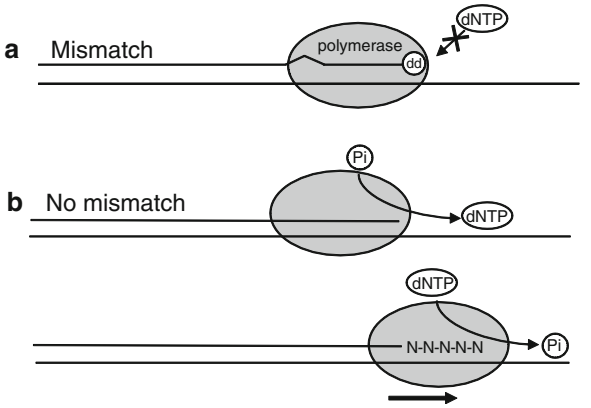


Fig. 1. (**a**) In the presence of a mismatch between primer and genomic DNA, the polymerase is unable to remove the blocking nucleotide (dd), and no polymerisation occurs. (**b**) If the primer is properly annealed, the polymerase is able to remove the blocking nucleotide in the presence of pyrophosphate (Pi); this is followed by polymerisation.

is followed by product capture using magnetic beads, linker ligation, amplification, and labelling. The labelling step provides another opportunity to use a blocked primer especially for allele-specific PAP-LMPCR. This can be done on material produced either by conventional LMPCR or PAP-LMPCR to improve the specificity of the labelling reaction so that only the desired sequence is labelled. Finally, the labelled products are visualised on a gel or capillary sequencer. A schematic representation of the procedure is shown in Fig. 2.

2. Materials

2.1. Primer Extension

1. 5′ Biotinylated, 3′ ddNTP blocked first primer 20 µM in 0.1× TE, stored at –20°C in small aliquots (Integrated DNA Technologies).
2. 2 U/µl Therminator™ DNA polymerase (New England Biolabs), stored at –20°C.
3. 10× Primer extension buffer: 200 mM Tris–HCl pH 8, 100 mM KCl, 100 mM $(NH_4)SO_4$, 30 mM $MgCl_2$, 1% Triton, 600 µM Na pyrophosphate, stored at –20°C.
4. Sulpholan (tetramethylene sulfone; Sigma), stored at room temperature.
5. 20 mM Tetramethylammonium oxalate (TMA ox; Sachem Inc), stored at –20°C.
6. dNTP mix (25 mM each stored at –20°C in small aliquots).

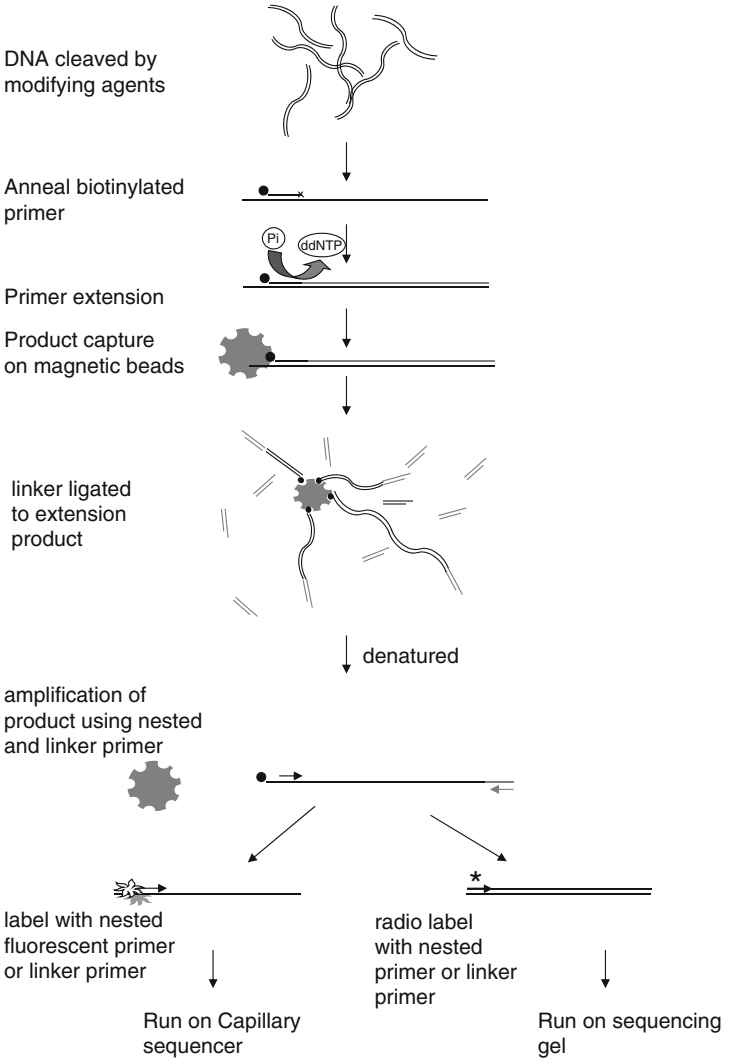


Fig. 2. *Schematic representation of PAP-LM-PCR.* The DNA is fragmented by DNA modifying agents such as DMS, piperidine, DNaseI, or MNase. When the blocked biotinylated primer is correctly annealed, pyrophosphorolysis removes the blocking nucleotide, and primer extension occurs. The primer extension products are captured on magnetic beads, and all other DNA molecules are washed away. This is followed by the ligation of a linker to provide a universal priming site for the amplification step. Amplification is followed by labelling the product either fluorescently or radioactively and then running it on the appropriate gel system.

2.2. Product Capture

1. 10 mg/ml Streptavidin beads (Dynal M-280), stored at 4°C; do not freeze.
2. 2× Binding and washing buffer (BW buffer): 10 mM Tris–HCl, pH 7.5, 1 mM EDTA, 2.0 M NaCl.
3. Magnetic separator.

2.3. Linker Ligation

1. LP25 oligonucleotide (GCGGTGACCCGGGAGATCTGA-ATTC; 200 μM in 1× TE).
2. LP21 oligonucleotide (GAATTCAGATCTCCCGGGTCA); 200 μM in 1× TE.
3. 10 U/μl T4 DNA ligase (Epicentre technologies), stored at –20°C.
4. T4 DNA Ligase 10× reaction buffer: 330 mM Tris–acetate (pH 7.8), 660 mM potassium acetate, 100 mM magnesium acetate, 5 mM DTT. Store at –20°C.
5. 50% Polyethylene glycol (PEG) 6000, in water.
6. 25 mM ATP, stored at –20°C in small aliquots.

2.4. Extension Product Amplification

1. Pfu turbo DNA polymerase (Stratagene) stored at –20°C.
2. 10× Cloned Pfu buffer: 200 mM Tris–HCl, pH 8.8, 20 mM $MgSO_4$, 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 1% Triton® X-100, 1 mg/ml nuclease-free bovine serum albumin (BSA), stored at –20°C.
3. Second primer: 20 μM in 0.1× TE, stored at –20°C in small aliquots.
4. LP25 oligonucleotide (GCGGTGACCCGGGAGATCTGA-ATTC; 20 μM in 0.1× TE).
5. 5 M Betaine (in water), stored at –20°C.
6. dNTP mix (25 mM each) stored at –20°C in small aliquots.

2.5. Radioactive Labelling

1. 10× PNK buffer: 700 mM Tris–HCl, pH 7.6, 100 mM $MgCl_2$, 50 mM DTT. Store at –20°C.
2. 10 U/μl T4 polynucleotide kinase (T4-PNK; New England Biolabs). Store at –20°C.
3. Third primer: 20 μM in 0.1× TE, stored at –20°C in small aliquots.
4. 370 MBq/ml Redivue™ [γ-32P] ATP (~110 TBq/mmol; Amersham Biosciences).
5. Pfu turbo DNA polymerase (Stratagene), stored at –20°C.
6. 10× Cloned Pfu buffer: 200 mM Tris–HCl, pH 8.8, 20 mM $MgSO_4$, 100 mM KCl.
7. 100 mM $(NH_4)_2SO_4$, 1% Triton® X-100, 1 mg/ml nuclease-free BSA, stored at –20°C.
8. 5 M Betaine (in water), stored at –20°C.
9. dNTP mix (2.5 mM each stored at –20°C in small aliquots).
10. MicroSpin™G-25 column (Amersham Biosciences). (Spin column containing Sephadex G25 DNA Grade F equilibrated in water with 0.05% Katon™.)

2.6. Sample Analysis on Denaturing Acrylamide Gels

1. 8 M Urea – 6% acrylamide (acrylamide/bis, 19/1) in 1× TBE.
2. 1× TBE (89 mM Tris–borate, 2 mM EDTA). For the stock solution, i.e. 5×, autoclaving is recommended to avoid formation of precipitates during storage.
3. 25% Ammonium persulfate (in water), stored at –20°C. Once thawed, do not refreeze.
4. TEMED (*N*,*N*,*N*′,*N*′-tetramethylethylenediamine; Sigma), stored at 4°C.
5. Formamide–EDTA–XC–BPB gel loading buffer: 10 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol FF in formamide.
6. 100% Ethanol.
7. Gel fixation buffer: 10% methanol, 10% acetic acid in water.

2.7. Non-radioactive Labelling

1. Pfu turbo DNA polymerase (Stratagene), stored at –20°C.
2. 10× Cloned Pfu buffer: 200 mM Tris–HCl, pH 8.8, 20 mM $MgSO_4$, 100 mM KCl, 100 mM $(NH4)_2SO_4$, 1% Triton® X-100, 1 mg/ml nuclease-free BSA, stored at –20°C.
3. Sulpholan (tetramethylene sulfone, Sigma), stored at room temperature.
4. dNTP mix (2.5 mM each), stored at –20°C in small aliquots.
5. 5′ Fluorescently labelled primer (Proligo), store at –20°C in the dark (1 pM/µl).
6. Cleanseq (Beckmann).
7. 85% Ethanol.

2.8. Capillary Electrophoresis

1. Capillary sequencing machine (in our case a CEQ 8000 from Beckman).
2. Fluorescently labelled marker for capillary sequencer.
3. Sample loading solution (Beckman).

2.9. Allele-Specific PAP-LMPCR

PAP labelling mix: 50 mM Tris–HCl pH 7.8, 16 mM $(NH_4)_2SO_4$, 3.5 mM $MgCl_2$, 100 µM dATP, dCTP, and dTTP, 100 µM 7 deaza dGTP, 90 µM Na pyrophosphate, 0.1 µM 3′ dideoxy-terminated 5′-labelled allele-specific primer (see Note 4), 1× Q solution (Qiagen), 5 U/reaction Klentaq (Ab Peptides Inc) 2 U/reaction Qiagen Taq DNA polymerase.

3. Methods

3.1. PAP Initiated LMPCR

3.1.1. Primers

The first primer should be 5′ biotinylated and 3′ dideoxyterminated and preferably should be HPLC-purified, since purity of this oligonuceotide is crucial (see Note 5). The second nested primer can partially overlap the first or be a short distance away. A third

nested primer or the linker primer can be used for labelling (see Note 6).

3.1.2. Primer Extension

In this crucial step, the first biotinylated primer is extended until it reaches a break point in the DNA but only if the blocking nucleotide is removed by pyrophosphorolysis. The amount of piperidine-treated DNA used can be varied, but 1 μg is a good starting point (see Note 7).

3.1.2.1. Primer Extension Reaction

1. In 0.2-ml thin-wall PCR tubes mix: DNA 1 μg, 10× Primer extension buffer 0.5 μl, (20 mM TMA ox 0.5 μl, optional), (see Note 8) 10.5 M sulpholan 0.14 μl, 25 mM dNTP mix 0.05 μl, 20 mM primer 0.05 μl, Therminator™ polymerase 1.00 μl, dH_2O to 5.00 μl.
2. Cover the reaction with 5–10 μl mineral oil (see Note 9).
3. Place in a thermal cycler with the following program: denaturation 95°C 20 min, primer-specific annealing temperature 20 min, (see Note 10) pyrophosphorolysis 60°C 2 min + 68°C 5 min, elongation 72°C 20 min (see Note 11).

3.1.3. Primer Extension Product Capture

The primer extension products are captured on magnetic beads, and all other DNA molecules are removed by washing.

1. Wash 15 μl Dynal M280 beads (10 mg/ml) three times with a large volume of 2× BW buffer and separate them with a magnetic separator after each wash.
2. Resuspend the beads in 5 μl 2× BW buffer and add to the primer extension mix. Incubate with rotation for 30 min to 1 h to capture the primer extension product (see Note 12).
3. Add 50 μl TE to the sample to aid separation of the beads preventing bead loss (see Note 13).
4. Wash the beads twice with 50–100 μl 1× TE and separate them with a magnetic separator after each wash.
5. Wash once with 50–100 μl 1 mM Tris–HCL pH 7.5.
6. Separate on a magnetic separator and keep the beads.

3.1.4. Ligation

A linker is added to the primer extension products on the beads to add a second, generic priming site for PCR amplification (see Note 14).

1. To make the linker, add 10 μl LP25 (200 μM), 10 μl LP21 (200 μM), and 80 μl dH_2O.
2. Heat to 95°C and cool slowly; this can be done in a thermal cycler [95°C, 3 min; 65°C, 15 min (slope 0.01°C/s); 55°C, 15 min (slope 0.01°C/s); 45°C, 15 min (slope 0.01°C/s); 37°C, 15 min (slope 0.01°C/s); 25°C, 15 min (slope 0.01°C/s)].

3. For the linker ligation reaction mix: 10× ligase buffer supplied 1.5 µl, 50% PEG 6000 3.0 µl, (see Note 15) Linker 3.0 µl, 25 mM ATP 0.6 µl, Epicentre T4 ligase (10 U/µl) 0.5 µl, H_2O to15 µl.
4. Resuspend the beads in the 15 µl ligation mix and incubate overnight at 16°C (see Note 16).

3.1.5. Amplification

The captured primer extension product is amplified using a nested specific primer (see Note primer) and the LP25 linker primer.

1. Add 50 µl TE to the ligation reaction before separation.
2. Separate the beads on a magnetic separator and remove supernatant.
3. Resuspend beads in 50 µl TE.
4. Separate the beads on a magnet and remove supernatant.
5. Repeat washing once with TE and then with 1 mM Tris–HCI pH 7.5.
6. Resuspend beads in 10 µl 1 mM Tris–HCI pH 7.5.
7. Heat the beads to 95°C for 15 min.
8. Cool on ice.
9. Mix: 10× Pfu buffer 5.0 µl, 5 M betaine 14 µl, 25 mM dNTPs 0.5 µl, second primer (20 mM) 0.5 µl, LP25 (20 mM) 0.5 µl, Pfu Turbo (2.5 U/µl) 1.0 µl, dH_2O 8.5 µl, and 10 µl bead suspension (final volume 50 µl).
10. Place in a thermal cycler with following program: denaturation 95°C 5 min [denaturation 95°C 45 s, specific annealing temperature 3 min, extension 72°C, 5 min] 22–25 cycles, completion 72°C 10 min.

3.1.6. Labelling of Reaction Products

To visualise the LM-PCR result, part of the amplification product needs to be labelled using either a radioactive or a fluorescent LP25 primer. Alternatively, a gene-specific nested third primer can be used. Radioactive samples are run on a sequencing gel, while fluorescently labelled samples can be run on a capillary sequencer. Use of a capillary sequencer enables the user to avoid the use of radiation; it is significantly quicker and more convenient than sequencing gels and allows the user to perform further analysis of the results (15). A primer labelled with a fluorophore suitable for the equipment upon which the PAP-LMPCR is to be run must be acquired. The experiments detailed here were run on a Beckman Coulter CEQ capillary sequencer, but we have also run fluorescently labelled reactions on ABI and LiCOR instruments; it is likely that other sequencers can also be used (see Note 17).

3.1.6.1. Radioactive Labelling

1. Radiolabel a primer by mixing: 10× PNK buffer 1 µl, dH_2O 1 µl, T4 Kinase (10 U/µl) 1 µl, 20 µM primer 2 µl, $\gamma^{32}P$ ATP 5 µl.
2. Incubate at 37°C for 60 min and then add 40 µl of dH_2O and remove unincorporated ^{32}P with a MicroSpin™G-25 column.
3. Set up a labelling reaction using the radiolabelled primer by mixing: 10× Pfu buffer 0.4 µl, 5 M betaine 1 µl, 2.5 mM dNTPs 0.4 µl, Pfu turbo (0.5 U/µl) 0.2 µl, radiolabelled primer 2.0 µl, amplification product 5–10 µl.
4. Place in a thermal cycler with following program: denaturation 95°C 5 min [denaturation 95°C 45 s, Specific annealing temperature 3 min, extension 72°C, 5 min] 6–7 cycles, completion 72°C 10 min (see Note 18).

3.1.6.2. Fluorescent Labelling

1. For the labelling reaction mix – 10× Pfu buffer – 0.4 µl, 10.5 M sulpholan – 0.11 µl, 2.5 mM dNTPs – 0.4 µl, Pfu turbo (0.5 U/µl) – 0.2 µl, fluorescently labelled primer (1 pM/µl) – 1.0 µl, dH_2O – 1.9 µl.
2. Place in a thermal cycler with following program; denaturation – 95°C 5 min [denaturation 95°C 45 s, Specific annealing temperature – 3 min, Extension – 72°C, 5 min] 7–8 cycles, completion – 72°C 10 min (see Note 19).

3.1.7. Electrophoresis of Radiolabelled DNA on a Polyacrylamide DNA-Sequencing Gel

1. Prepare the sequencing gel with 0.4-mm spacers and comb (6% polyacrylamide, 8 M urea, acrylamide/bisacrylamide, 19/1) (see Note 20).
2. Assemble the gel apparatus and pre-run with 1× TBE for more than 60 min and confirm that the gel temperature is greater than 50°C.
3. Add formamide loading buffer to the labelled PCR products (1/1) and denature the DNA by incubation for 5 min at 100°C and quickly chill on ice.
4. After the pre-run, flush the urea from the wells using a syringe and needle.
5. Load the denatured samples (using flat-ended gel loading tips) into the cleaned wells.
6. Electrophoresis is performed at 80–90 W (for a 40 cm × 40 cm × 0.4 mm gel). If a long and a short run are performed on one plate, the second series of samples should be loaded when the dye front has reached approximately 10 cm from the bottom of the gel. BPB has, approximately, the same mobility as the primers, so stop the electrophoresis when the BPB has just run out from the gel.
7. Separate the glass plates and fix the gel with fixation buffer for approximately 10 min. Blot the gel onto 3-MM Whatman

paper, cover the gel with Saran Wrap and dry the gel on a gel dryer at 80°C.

8. When the gel is dried, expose overnight to X-ray film or a Phospho Imager screen.

3.1.8. Electrophoresis of Fluorescently Labelled DNA on a Capillary-Sequencer

It is necessary to remove the unincorporated primer and salts before loading onto a capillary sequencer. We find CleanSeq (Beckman Coulter) used as per the manufacturer's instructions works well.

Settings for the Beckman CEQ are as follows (settings for other instruments may vary): Injection 2.5 kV 30 s; denaturation off; capillary temperature 45°C; wait for capillary temperature yes; separation 6 kV for 35 min. Analysis parameters: slope threshold 8%; relative peak height threshold 2%. All other settings were left at default.

3.2. Allele-Specific PAP-LMPCR

In allele-specific PAP-LMPCR the steps previous to labelling were carried out as in (16) except the primer extension oligo was 5′ biotinylated and was captured and ligated as in Subheadings 3.1.3–3.1.4. In the amplification step, the Pfu turbo and betaine were replaced with Qiagen Taq and Q solution. Whichever polymerase and buffer system are used in the amplification step, it is important that the final pH of the labelling reaction is pH 8. The general principle is outlined in Fig. 3.

1. For the labelling reaction mix: 10 µl amplification reaction, 20 µl PAP labelling mix (giving a final composition of; 36.3 mM Tris–HCl, 14 mM $(NH_4)_2SO_4$, 60 µM Na Pyrophosphate, 16.5 mM KCl, 3 mM $MgSO_4$, 100 µM dNTPs, pH 8).
2. Place in a thermal cycler with following program: [denaturation 95°C 2 min, pyrophosphorolysis 60°C 2 min, 68°C

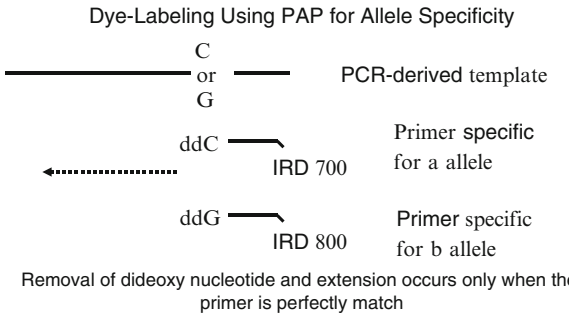


Fig. 3. *Allele specific PAP-LMPCR*. Schematic diagram of allele-specific PAP-LMPCR showing two primers labelled for LiCOR, one for each allele. Each primer will only extend when bound to its corresponding allele; the resulting products are then visualised on the separate channels of the LiCOR. For other systems such as capillary sequencers, two separate reactions would be required which are then run separately.

1.5 min, extension 72°C 2.5 min] 1 cycle [denaturation 95°C 45 s, pyrophosphorolysis 60°C 2 min, 68°C 1.5 min, extension 72°C 2.5 min] 15 cycles.

3.3. Representative Examples of In Vivo Footprinting Experiments Involving PAP-LMPCR

An example of PAP-LMPCR using DNaseI-treated DNA that was subsequently fluorescently labelled using a labelled linker primer and run on a capillary sequencer is shown in Fig. 4. The reaction was carried out on the promoter of the *Csf1r* gene. *Csf1r* is the gene encoding for the receptor for colony stimulating factor one and is an essential gene for macrophage development. The sequence shown contains a run of 20 T's, while in the region upstream through which the reaction passed to get to this point is a 24 CA dinucleotide repeat (not shown). The PAP-LMPCR reaction has successfully passed through both repeats which conventional LM-PCR is unable to do. In previous DNaseI LM-PCR studies of this region which analysed the opposite strand, it was shown that the chromatin remains largely open in B cells (17) even though the gene is not expressed. Subsequently, it was been shown that the PU.1 site at –130 bp is important for *Csf1r* activation in macrophages, while the binding of PAX5 over the transcription start site is required for silencing in B cells (18). The experiment shown here confirms that the chromatin accessibility is broadly the same in both macrophages and B cells. However, there are differences around the transcription start site; in the splenic B cells, the accessibility of both the PU.1 site at –130 bp and the adjacent transcription start occupied by PAX5 is reduced

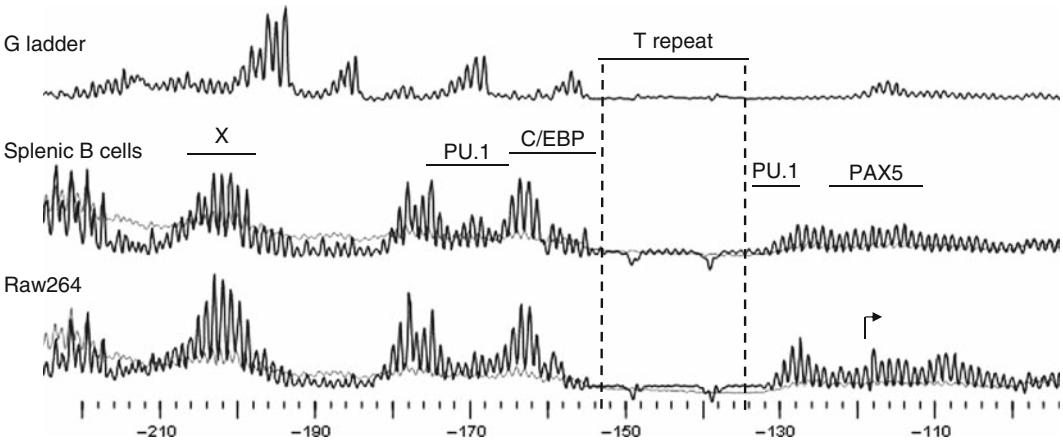


Fig. 4. *DNaseI PAP-LMPCR on the Csf1r gene in macrophages and B cells. Csf1r* is the colony-stimulating receptor gene, which is expressed in macrophages, but not in B cells. Previous DNaseI accessibility studies of the opposite strand have shown that the chromatin remains open in B cells (17); however, prior to PAP-LMPCR, we were unable to examine both strands. Subsequently, it was shown that the PU.1 site at –130 bp is important for *Csf1r* activation in macrophages, while the binding of PAX5 over the transcription start site is required for silencing in B cells (18). The experiment shown here confirms that the chromatin accessibility is broadly the same in both macrophages and B cells.

as compared to macrophages. This is further supporting evidence that the silencing of the *Csf1r* gene in B cells is not due to major changes in chromatin structure but is carried out by the interaction of PAX5 with PU.1 and the basal transcriptional machinery.

4. Notes

1. DMS and piperidine are highly toxic. Perform all manipulations in fume hoods. All tips used with DMS should be rinsed in 1-M NaOH, and all waste solutions containing DMS should be treated with NaOH before disposal to inactivate the DMS.
2. The amount of product produced in the primer extension reaction is less than that in conventional LMPCR, as any incorrectly bound primer will not elongate and some of the primer that is correctly bound will not undergo pyrophosphorolysis. This can be compensated for by increasing the amount of DNA in the starting reaction and/or adding an extra cycle or two to the amplification or labelling reaction. It is inadvisable to add a large number of cycles, as this can lead to artefacts in the final gel.
3. Therminator™ DNA Polymerase is a 9°N™ DNA Polymerase variant with an enhanced ability to incorporate modified substrates such as dideoxynucleotides, ribonucleotides, and acyclonucleotides. New variants with enhanced ability to handle modified nucleotides are now available but have not been tried in PAP-LMPCR.
4. When designing allele-specific primers, the last base (3′) of the primer should correspond to a known SNP so that only the primer matching to the allele of interest will extend (see Fig. 4). The primer should be labelled according to the system upon which it is going to be run. If using a LiCOR system, then a primer corresponding to each allele can be included in the reaction, one labelled with IRD700 and the other with IRD800. Both alleles will be labelled in the reaction but visualised separately on the LiCOR. For radiolabelling and most other gel systems, two separate labelling reactions are needed, one for each allele.
5. The presence of any oligo nucleotide without the blocking nucleotide will lead to polymerisation without pyrophosphorolysis; this may lead to higher background from non-specific hybridisation.
6. All enzymes and reaction buffers are stored at –20°C. ATP, DTT, and dNTPs are very sensitive to repeated freeze-thaw cycles. Therefore, they must be stored frozen in small aliquots.

Oligonucleotides are especially unstable in diluted solution. To minimise any degradation, they must be reconstituted with 1× TE at a concentration of 200 μM or more and stored at −20 or −80°C in small aliquots.

7. The required amount of DNA is totally dependent on the performance of the primer sets and sequence of the target region. It has to be tested for each primer set.
8. There are a wide range of chemicals reported to be PCR enhancers; the most useful ones in this procedure have been found to be sulpholan (19), DMSO, and betaine (20). Betaine monohydrate (*N*,*N*,*N*-trimethylglycine) is a zwitterionic osmoprotectant found to alter DNA stability such that G/C-rich regions melt at temperatures more similar to A/T-rich regions and eliminate pausing by DNA polymerases. It is not absolutely necessary to have this component in the PCR reaction. However, we observed reproducible and significant improvement in the efficiency and accuracy of the amplification by adding betaine. The use of TMA (21) in the primer extension reaction sometimes is beneficial. The use of sulpholan is usually very beneficial, as it seems much better at helping to make difficult sequences amenable to PCR. In some cases, 5% DMSO may still be the enhancer of choice.
9. Mineral oil should still be used even with a heated lid to prevent evaporation as the reaction volume is very small.
10. It is often beneficial to try and optimise the annealing temperature of the primer using a gradient PCR machine, as T_m is often different from that predicted. Therefore, determine the best annealing temperature by performing normal PCR using the test primer and one on the opposite strand (with T_d 5° higher than the primer to be tested). Then carry out PCR in increasing steps of 2–3°C across the T_m for 50 mM NaCl. The highest temperature at which efficient PCR still occurs is defined as T_m, and use T_m − 2°C as the annealing temperature for the PAP-LMPCR.
11. The primer extension temperature can be raised from 72 to 76°C if the sequence is very GC rich, since the polymerase is fully stable at this temperature.
12. All primers, whether extended or not, will be captured; however, as the amount of primer is very small it will not interfere with subsequent PCR steps.
13. Increasing the volume allows the beads to move against the side of the tube easier, aiding the removal of the supernatant without inadvertently sucking up the beads.
14. The product capture and ligation steps are switched around from conventional PCR as the presence of the pyrophosphate containing buffer may inhibit ligation of the linker.

15. Primer design: the lengths of the primers are normally between 21 and 25 bp. Generally, longer primers perform better. The most critical primer is the first primer. Primers must be nested, but it is not important whether or not the first and second primers overlap. However, the second and third primers ideally should overlap by at least 15 bases to avoid interference between primers. Ideally, the T_d of the primer should increase in the order $1 < 2 < 3$ with ideal T_{ds} being in the regions of 52 ± 6 for primer 1, 70 ± 3 for primer 2, and 77 ± 5 for primer 3.
16. PEG has been reported to enhance the ligation efficiency (22). However, the quality of PEG is extremely variable. Purchase a good-quality PEG and check it in a ligation reaction before use. In some cases, no significant differences were found in reactions either with or without PEG.
17. Most fragments ligate within 2 h, but some DNA ends ligate more efficiently than others. Leaving them overnight at 16°C will, therefore, help you to get more uniform band patterns. Blunt-end ligations work best at 16°C but can also work in the range of 4–22°C.
18. The number of cycles for the labelling can be changed depending on the amplification efficiency and amount of starting materials.
19. Sulpholan (0.3–0.4 M) can be used instead of betaine and may give a greater product yield; however, the presence of sulpholan in the labelling reaction reduces the resolution on the acrylamide gel, producing fuzzy bands. Sulpholan is recommended when using a capillary sequencer, as it is removed during sample preparation for loading.
20. An example for the procedure to prepare a polyacrylamide DNA-sequencing gel is as follows: (1) Prepare a 40% acrylamide solution by dissolving 380 g of acrylamide, 20 g of *N*,*N*-methylenebisacrylamide in 1 l of water and filtering the solution through a nitrocellulose 0.45-µm filter. Note that acrylamide powder is toxic and that there are number of commercially available ready-made solutions that are easier and safer to use. (2) Prepare a 6% acrylamide gel solution (75 ml of 40% acrylamide solution, 100 ml of 5× TBE, 240 g urea/500 ml) and filter through a nitrocellulose 0.45-µm filter. This final solution can also be purchased (e.g. SequaGel system from Flowgen), in which case, make up the solution according to the manufacturer's instructions. (3) Assemble the glass plates. Lay the larger plate flat on the bench, put spacers along the sides and place the smaller plate on top of it. (4) Add 50 µl of 25% ammonium persulphate and 50 µl of TEMED to 50 ml of 6% acrylamide gel solution. (5) Pour the

gel solution between two glass plates while tapping the top glass to avoid formation of bubbles. (6) Insert the comb and clamp the comb and glass plates using binder clips.

References

1. Montecino, M., Stein, J.L., Stein, G.S., Lian, J.B., van Wijnen, A.J., Cruzat, F., Gutierrez, S., Olate, J., Marcellini, S., and Gutierrez, J.L. (2007) Nucleosome organization and targeting of SWI/SNF chromatin-remodeling complexes: contributions of the DNA sequence. *Biochemistry and Cell Biology = Biochimie et Biologie Cellulaire* **85**, 419–25.
2. Racki, L.R. and Narlikar, G.J. (2008) ATP-dependent chromatin remodeling enzymes: two heads are not better, just different. *Current Opinion in Genetics & Development* **18**, 137–44.
3. Pfeifer, G.P. (2006) Measuring the formation and repair of DNA damage by ligation-mediated PCR. *Methods in Molecular Biology* **314**, 201–14.
4. Kontaraki, J., Chen, H.H., Riggs, A., and Bonifer, C. (2000) Chromatin fine structure profiles for a developmentally regulated gene: reorganization of the lysozyme locus before trans-activator binding and gene expression. *Genes & Development* **14**, 2106–22.
5. Tagoh, H., Himes, R., Clarke, D., Leenen, P.J., Riggs, A.D., Hume, D., and Bonifer, C. (2002) Transcription factor complex formation and chromatin fine structure alterations at the murine c-fms (CSF-1 receptor) locus during maturation of myeloid precursor cells. *Genes & Development* **16**, 1721–37.
6. Lefevre, P., Lacroix, C., Tagoh, H., Hoogenkamp, M., Melnik, S., Ingram, R., and Bonifer, C. (2005) Differentiation-dependent alterations in histone methylation and chromatin architecture at the inducible chicken lysozyme gene. *The Journal of Biological Chemistry* **280**, 27552–60.
7. Tagoh, H., Cockerill, P.N., and Bonifer, C. (2006) In vivo genomic footprinting using LM-PCR methods. *Methods in Molecular Biology* **325**, 285–314.
8. Liu, Q. and Sommer, S.S. (2000) Pyrophosphorolysis-activated polymerization (PAP): application to allele-specific amplification. *BioTechniques* **29**, 1072–76, 1078, 1080 passim.
9. Liu, Q. and Sommer, S.S. (2004) PAP: detection of ultra rare mutations depends on P* oligonucleotides: "sleeping beauties" awakened by the kiss of pyrophosphorolysis. *Human Mutation* **23**, 426–36.
10. Ingram, R., Gao, C., Lebon, J., Liu, Q., Mayoral, R.J., Sommer, S.S., Hoogenkamp, M., Riggs, A.D., and Bonifer, C. (2008) PAP-LMPCR for improved, allele-specific footprinting and automated chromatin fine structure analysis. *Nucleic Acids Research* **36**, e19.
11. Liu, Q. and Sommer, S.S. (2004) Pyrophosphorolysis by type II DNA polymerases: implications for pyrophosphorolysis-activated polymerization. *Analytical Biochemistry* **324**, 22–8.
12. Liu, Q. and Sommer, S.S. (2002) Pyrophosphorolysis-activatable oligonucleotides may facilitate detection of rare alleles, mutation scanning and analysis of chromatin structures. *Nucleic Acids Research* **30**, 598–604.
13. Gardner, A.F. and Jack, W.E. (1999) Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Research* **27**, 2545–53.
14. Gardner, A.F. and Jack, W.E. (2002) Acyclic and dideoxy terminator preferences denote divergent sugar recognition by archaeon and Taq DNA polymerases. *Nucleic Acids Research* **30**, 605–13.
15. Ingram, R., Tagoh, H., Riggs, A.D., and Bonifer, C. (2005) Rapid, solid-phase based automated analysis of chromatin structure and transcription factor occupancy in living eukaryotic cells. *Nucleic Acids Research* **33**, e1.
16. Pfeifer, G.P. and Dammann, R. (1999) Measuring the formation and repair of UV photoproducts by ligation-mediated PCR. *Methods in Molecular Biology* **113**, 213–26.
17. Tagoh, H., Schebesta, A., Lefevre, P., Wilson, N., Hume, D., Busslinger, M., and Bonifer, C. (2004) Epigenetic silencing of the c-fms locus during B-lymphopoiesis occurs in discrete steps and is reversible. *EMBO Journal* **23**, 4275–85.
18. Tagoh, H., Ingram, R., Wilson, N., Salvagiotto, G., Warren, A.J., Clarke, D., Busslinger, M., and Bonifer, C. (2006) The mechanism of repression of the myeloid-specific c-fms gene by Pax5 during B lineage restriction. *EMBO Journal* **25**, 1070–80.

19. Chakrabarti, R. and Schutt, C.E. (2001) The enhancement of PCR amplification by low molecular-weight sulfones. *Gene* **274**, 293–8.
20. Mytelka, D.S. and Chamberlin, M.J. (1996) Analysis and suppression of DNA polymerase pauses associated with a trinucleotide consensus. *Nucleic Acids Research* **24**, 2774–81.
21. Kovarova, M. and Draber, P. (2000) New specificity and yield enhancer of polymerase chain reactions. *Nucleic Acids Research* **28**, E70.
22. Pheiffer, B.H. and Zimmerman, S.B. (1983) Polymer-stimulated ligation: enhanced blunt- or cohesive-end ligation of DNA or deoxyribooligonucleotides by T4 DNA ligase in polymer solutions. *Nucleic Acids Research* **11**, 7853–71.

Chapter 13

The Many Faces of MLPA

Thomas Ohnesorg, Erin Turbitt, and Stefan J. White

Abstract

Multiplex Ligation-dependent Probe Amplification (MLPA) is a PCR-based technique that was developed for identifying deletions and duplications in genomic DNA. The simplicity and sensitivity of this approach has led to it being implemented in many laboratories around the world. Since the original publication, there have been several variants of MLPA described, allowing the quantitative analysis of mRNA transcript levels, CpG methylation, complex genomic regions, and DNaseI hypersensitive sites. This chapter outlines the basic MLPA protocol, describes the different modifications and applications that have been published, and discusses the critical points during each of the steps.

Key words: MLPA, Copy number variation, Deletion, Duplication, Mutation screening, Methylation, Expression analysis, DNaseI hypersensitivity

1. Introduction

For many years, the detection of deletions and duplications in genomic DNA was a problem for mutation screening. Techniques such as FISH, Southern blotting, and qPCR were available but often avoided as they were time-consuming, expensive, and/or difficult to establish. The publication of Multiplex Ligation-dependent Probe Amplification (MLPA) (1) has dramatically changed this situation. MLPA is based on the ligation of two half-probes after they have annealed to a specific genomic sequence (Fig. 1). All probes within a mix have identical sequences at the ends, allowing ligated half-probes to be PCR-amplified with a single primer pair (one of which is fluorescently labeled). The resulting products are typically separated by capillary electrophoresis, and individual probes can be recognized on the basis of their unique length. Compared to other techniques, the principle advantages of MLPA are the simplicity of the reaction, the low amount of DNA required, the number of loci

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_13, © Springer Science+Business Media, LLC 2011

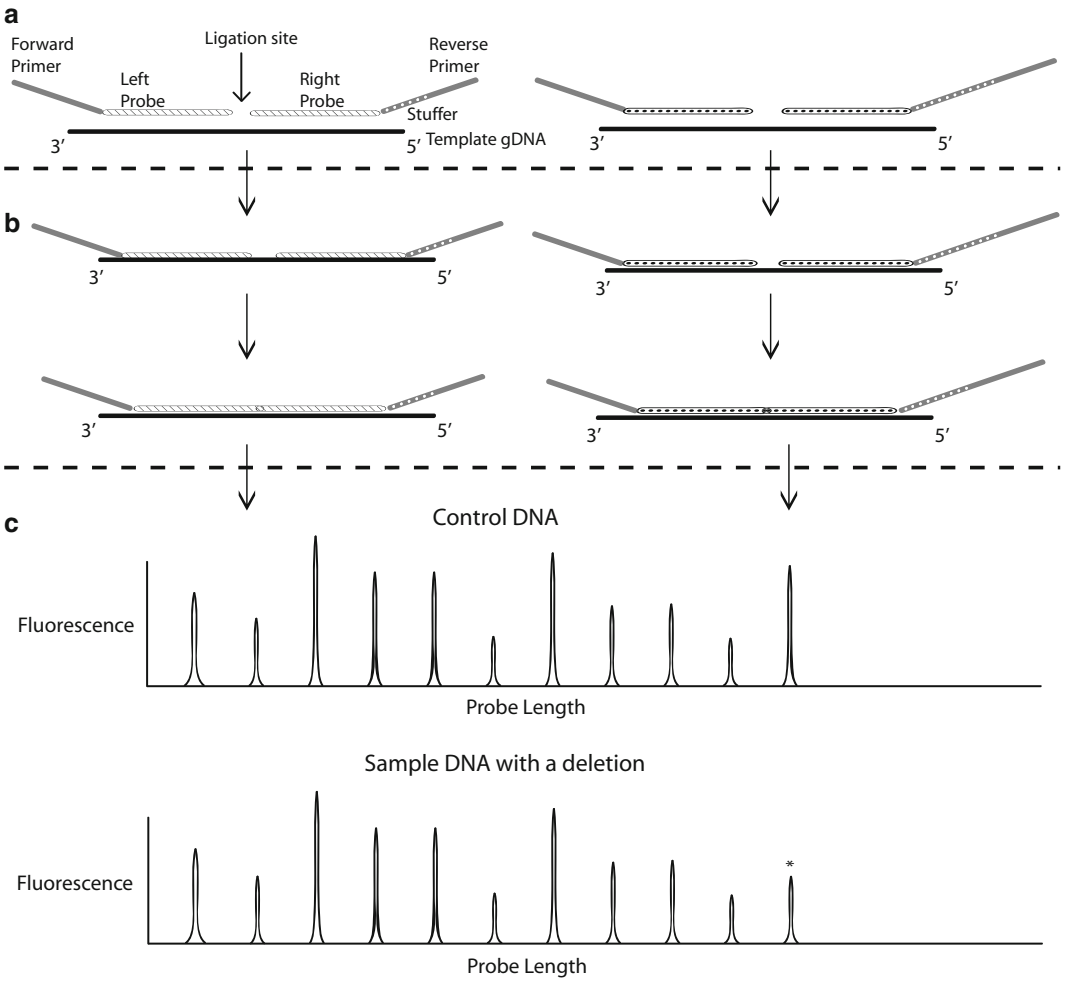


Fig. 1. Standard MLPA procedure. (**a**) Probes are designed to be complementary to the target region of interest. Multiple loci can be targeted in one reaction, with different probes having identical ends to allow PCR amplification with a single primer pair. Different stuffer sequences allow each probe to have a unique length. (**b**) The probes are hybridized to the genomic DNA and ligated together. Only ligated probes can be amplified. (**c**) The amplification products are separated using capillary electrophoresis. The peak marked with an *asterisk* corresponds to a deletion in the sample DNA.

that can be analyzed simultaneously, and the sensitivity in detecting copy number changes.

A complicating factor of MLPA is probe development. In the original description, the left-hand half-probes were synthetic oligonucleotides of 40–50 nt in length, whereas the right-hand half-probes were created by a cloning process involving modified M13 vectors. This allowed single-stranded oligonucleotides of up to 500 nt to be generated, but is a time-consuming and costly process. For assays that will be performed on many samples, the expense is worthwhile, and hundreds of different probe mixes are commercially available. If only a few samples are to be analyzed, the cost and time involved may be prohibitive.

An alternative approach is to design the half-probes in such a way that both can be chemically synthesized. This greatly simplifies probe development, as no cloning steps are involved, although length constraints in oligonucleotide synthesis mean that fewer probes can be used in a single mix. A way to get around this limitation is by developing distinct probe mixes that have different PCR primer binding sequences. These can be combined and distinguished from each other by using a different fluorophore for each specific amplification primer (2, 3).

The simplicity and sensitivity of MLPA has led to several variants being developed. The quantitative nature of MLPA means that it can also be used for determining RNA transcript levels (4, 5). As the ligase used in MLPA will not ligate DNA molecules in the context of a DNA–RNA duplex, it is first necessary to create a cDNA template for the MLPA oligonucleotides to bind to (Fig. 2a). This is done using a specific primer sequence that anneals next to each MLPA probe binding site.

Another application of MLPA is the analysis of CpG methylation in genomic DNA (6). The protocol takes advantage of a specific property of methylation-specific restriction endonucleases, namely, the methylation state of CpG dinucleotides within the specific restriction endonuclease recognition sequence determines whether or not digestion occurs (Fig. 2b). If the CpG is methylated, then the recognition sequence is insensitive to the restriction endonuclease, whereas a recognition sequence that contains an unmethylated CpG can be digested. Each MLPA probe is designed to hybridize to a genomic sequence that includes a recognition site for a methylation-specific restriction endonuclease, with four base-cutters such as HhaI giving the most flexibility.

Extension-MLPA was developed as a way to analyze genomic sequences that are not suitable templates for MLPA probe design (7). Examples include regions with extremely high or low GC%, palindromic sequences, and repetitive tracts. The unique feature of this approach is that the oligonucleotides that make up each MLPA probe do not anneal immediately adjacent to each other (Fig. 2c). Instead, the probes are designed in such a way that there is a gap between them (up to 114 bp has been reported). Prior to the oligonucleotides being ligated together, it is necessary for the left-hand probe to be extended until it reaches the right-hand probe. This is achieved by adding Stoffel fragment in addition to the DNA ligase. Stoffel fragment is a modified DNA polymerase lacking 5′ to 3′ exonuclease activity, therefore does not degrade the bound right-hand probe.

We have showed recently that MLPA can be employed to analyze and identify DNaseI hypersensitive sites (8), which are sites of open chromatin characteristic of regulatory elements such as enhancers and promoters. Such genomic regions are susceptible

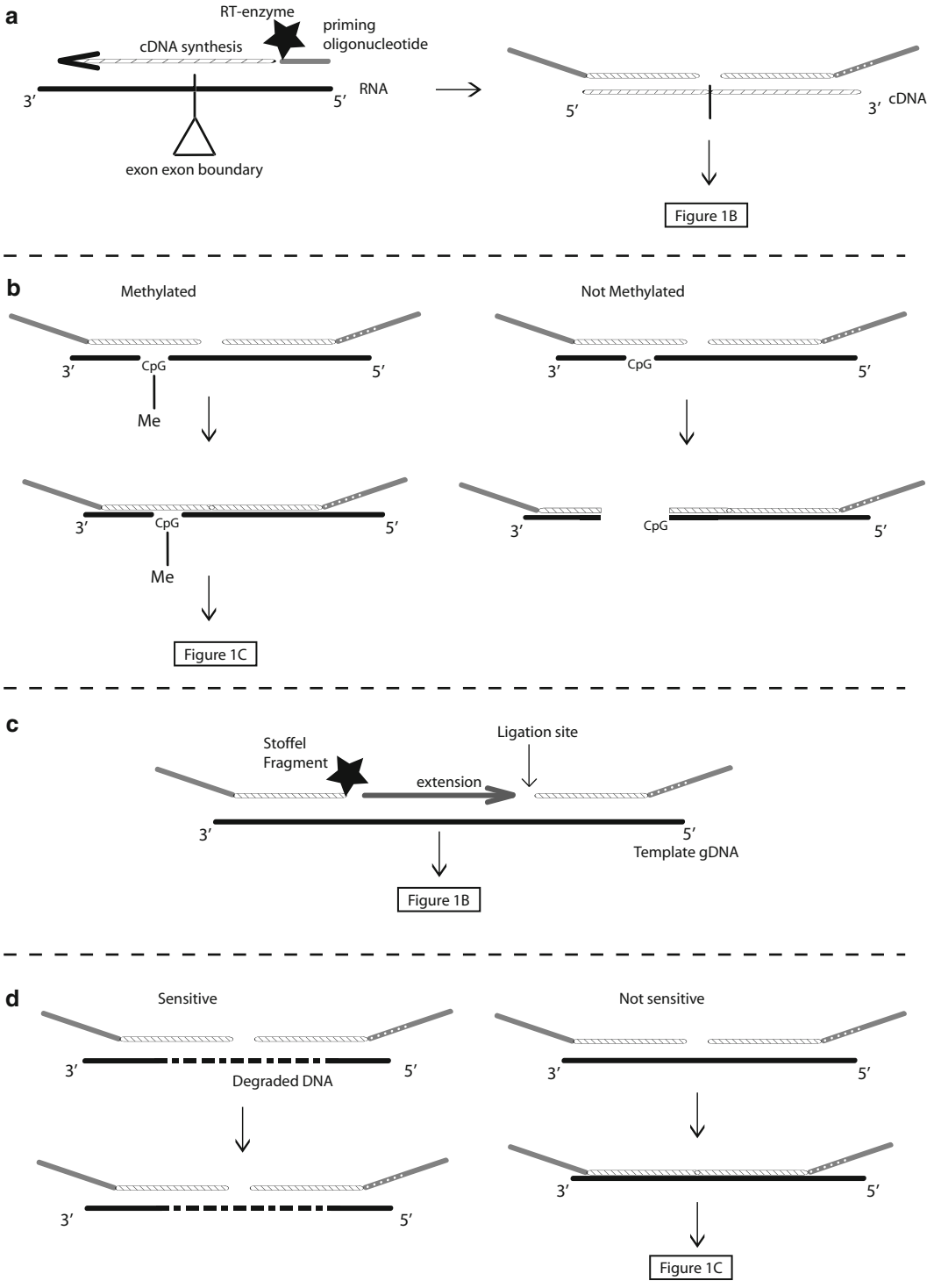


Fig. 2. Variations on the standard MLPA technique. (**a**) RT-MLPA (mRNA expression analysis). Target mRNA is reverse-transcribed into cDNA using specific primers. This cDNA is the template for probe hybridization and ligation. (**b**) Methylation MLPA. Following probe hybridization, DNA ligase and a methylation-sensitive restriction endonuclease are added. Unmethylated DNA will be digested, meaning that probe amplification cannot occur. Methylated DNA prevents digestion

to digestion by DNaseI and can be recognized in an MLPA reaction by the lack of signal from a specific probe (Fig. 2d). Depending on the loci of interest (many regulatory regions such as promoters have a high GC content, which may complicate probe design), it may be necessary to combine this approach with extension MLPA. This also makes it possible to cover larger genomic regions with a single probe, and standard MLPA probes can be combined with extension MLPA probes in a single mix.

2. Materials

2.1. Basic MLPA Reaction

1. All reagents for the basic MLPA protocol can be purchased from MRC-Holland, the Netherlands (http:\\www.mlpa.com). Each item can be recognized by a cap of a specific color.

   MLPA buffer (yellow cap).

   Ligase-65 (green cap: NB this used to be a brown cap in previous versions).

   Ligase-65 buffer A (transparent cap).

   Ligase-65 buffer B (white cap).

   SALSA enzyme dilution buffer (blue cap).

   SALSA PCR primers + dNTPs (brown cap: NB this used to be a purple cap in previous versions). The sequence of the primers are:

   – Forward 5′-GGGTTCCCTAAGGGTTGGA-3′ (fluorescently labeled at the 5′ end).

   – Reverse 5′-GTGCCAGCAAGATCCAATCTAGA-3′.

   SALSA DNA Polymerase (orange cap).
   SALSA probe mix (black cap).

   – It is also possible to develop "homemade" probes using synthetic oligonucleotides only. Criteria for MLPA probe development are described in (1) and are also available on the MRC-Holland website (http://www.mlpa.com). There are several important points to consider. Both left- and right-hand oligonucleotides should be targeted to unique sequences. If this is not possible, then a single mismatch at the ligation site of the two half-probes

Fig. 2. (continued) of the genomic DNA–MLPA probe hybrid, allowing subsequent probe amplification. (**c**) Extension MLPA. Half-probes do not anneal adjacently to each other but are designed to hybridize with up to 100 nt between them. An extension step using Stoffel fragment fills in this gap, allowing ligation. (**d**) DNaseI MLPA. Isolated nuclei are treated with DNaseI. DNaseI sensitive loci are degraded, preventing probe hybridization. Loci that are not sensitive will not degrade and can serve as templates for probe hybridization.

should theoretically be sufficient to distinguish the target sequence from the homologous sequence, although this needs to be tested experimentally for each sequence. The GC% of the hybridizing sequence of each half probe should be 35–60%, and the T_m should be greater than 66°C. To maximize signal strength, it has been shown that the first nucleotide of the unique sequence of the left-hand half-probe should be a G or C (1). To facilitate ligation, the right-hand half-probe should be phosphorylated at the 5′ end.

2.1.1. Additional Reagents

Hi-Di formamide (Applied Biosystems).
Size standard (Applied Biosystems).

2.2. Extension MLPA

1. Stoffel fragment (Applied Biosystems).
2. Nucleotides (10 mM each of dATP, dCTP, dGTP, and dTTP).

2.3. RT MLPA

1. MMLV reverse transcriptase enzyme (Promega).

2.4. DNaseI MLPA

1. RQ1 DNaseI (1 U/µl; Promega).
2. NE-PER Nuclear and Cytoplasmic Extraction kit (Thermo Scientific).
3. Nuclei lysis buffer (200 mM NaCl, 150 mM Tris–HCl pH 8, 10 mM EDTA pH 8, and 0.2% SDS).
4. Proteinase K.
5. RNaseA (Sigma).
6. HighPure PCR Purification Kit (Roche).

2.5. Methylation MLPA

1. HhaI (Promega).

3. Methods

3.1. Basic MLPA Protocol

1. The MLPA protocol below is based on that described in the original publication (1) and can also be found at http://www.mlpa.com
2. Add 20–500 ng genomic DNA in a final volume of 5 µl to a PCR tube (see Note 1). The DNA is denatured at 98°C for 5 min (see Note 2) and allowed to cool to room temperature for at least 5 min.
3. To the denatured DNA, add: 1.5 µl MLPA Probe-mix + 1.5 µl SALSA MLPA buffer. Mix with care (see Note 3). Incubate for 1 min at 95°C and then for 16 h at 60°C (see Note 4).

4. Ligase-65 mix (make at room temperature):

   Mix 3 µl Ligase-65 buffer A (transparent cap) + 3 µl Ligase-65 buffer B (white cap) + 25 µl H_2O. Add 1 µl Ligase-65 (green cap) and mix again.

   Reduce temperature of the thermal cycler to 54°C. *Keeping the PCR tubes in the thermal cycler*, add 32 µl Ligase-65 mix to each tube and mix. The reactions should be incubated for 10–15 min at 54°C and then heated for 5 min at 98°C to inactivate the ligase (see Note 5).

5. There have been different procedures described for preparing the PCR amplification, with the two most commonly used described below (see Note 6).

   (a) For each reaction prepare the following mix:

SALSA enzyme buffer (blue)	2 µl
SALSA PCR primers (brown)	2 µl
SALSA Polymerase (orange)	0.5 µl
H_2O	5.5 µl

   Combine 25 µl H_2O, 5 µl SALSA PCR buffer (red), and 10 µl MLPA ligation reaction in a PCR tube and place in a thermal cycler at 60°C.

   Add 10 µl of the previously prepared polymerase mix to each tube and start the PCR.

   (b) For each reaction:

SALSA PCR buffer (red)	5 µl
SALSA enzyme buffer (blue)	2 µl
SALSA PCR primers (brown)	2 µl
SALSA Polymerase (orange)	0.5 µl
H_2O	30.5 µl

   Aliquot 40 µl into PCR tubes on ice, then add 10 µl ligation mix.

6. The PCR reaction should be carried out with the following settings:
   1 cycle: 1 min 95°C
   35 cycles: 30 s 95°C, 30 s 60°C, 30 s 72°C
   1 cycle: 20 min 72°C.

7. Prepare the samples for fragment analysis (see Note 7).
   Add 5 µl size standard to 1 ml Hi Di formamide and mix well. Into each well of a 96-well, add 10 ml of the formamide/size standard mix.

   Add 1 µl of PCR product (see Note 8).

8. MLPA reactions are most commonly processed by fragment analysis on a capillary sequencer, although other possibilities that have been described include agarose gel electrophoresis

(9) and microarray hybridization (10). In a fragment analysis, DNA is separated by electrophoresis, with each probe represented by a peak of a defined length. The relative size (height or area) of each peak is used to estimate copy number. Different normalization approaches have been described based on the number of probes within the probe mix that may be affected. If the mix contains target loci spread across many chromosomes, it is unlikely that the majority of probes will be deleted or duplicated within a single sample. It is, therefore, possible to use the sum of all the probes when calculating probe ratios. In contrast, there may be circumstances when many probes may be deleted or duplicated (e.g., a probe mix that covers every exon of a gene). In these cases, it is better to normalize each peak against several control probes from diverse loci that should not be deleted or duplicated in any of the samples being screened.

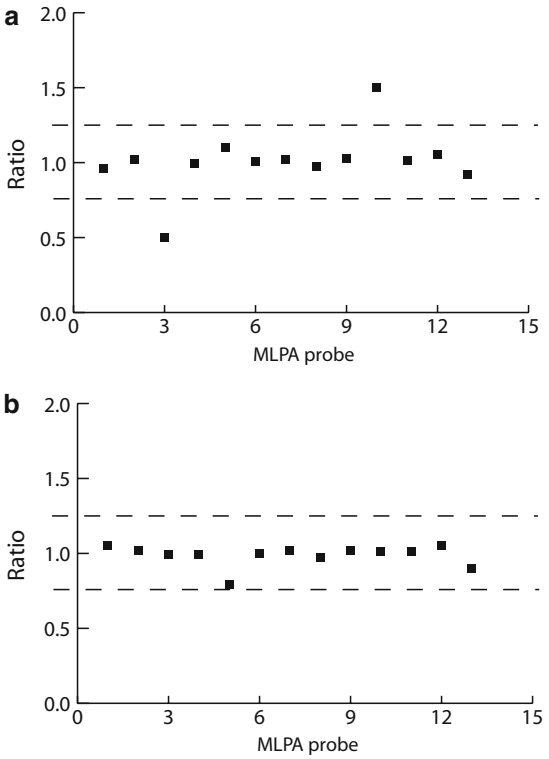


Fig. 3. MLPA data analysis. Following normalization, the ratios of probes representing unaffected loci will cluster around 1.0. (**a**) Deletions or duplications are identified by ratios significantly different from this, theoretically at 0.5 and 1.5, respectively. Arbitrary thresholds are often used, in this case indicated by the *horizontal dashed line*. (**b**) Intermediate ratios close to the thresholds, e.g., probe 5, may be due to noise, mosaicism, or sequence changes around the ligation site hindering probe annealing and subsequent ligation.

Once normalized ratios have been calculated, it is necessary to define thresholds for deletions and duplications. Assuming a normalized value of 1 corresponds to a copy number of 2, a deletion should be at 0.5 and a duplication should be at 1.5 (Fig. 3a). Thresholds approximately midway between these two points are commonly used (i.e., a deletion is below 0.75, and a duplication is above 1.25), although ratios close to the thresholds may be difficult to interpret without further investigation (Fig. 3b). One possibility is that the intermediate ratio seen is due to the rearrangement being present in only a subset of cells (mosaicism) (11).

Another possibility is a sequence change at or around the ligation site, which may disturb probe annealing sufficiently such that efficient ligation cannot take place. Any uncertainty can be resolved by designing a second probe in close proximity to the affected probe. If the reduced ratio is due to mosaicism, then the second probe should have a similar ratio to the first probe. If a sequence change around the ligation site of the first probe is the cause however, then the second probe should give a ratio around 1.0. Changes in sequence can be confirmed by direct sequencing.

For complex loci where the typical copy number is other than 2, the use of thresholds is not appropriate. In these situations, copy number can be estimated by clustering ratios into defined groups (12, 13).

3.2. RT-MLPA

1. Mix on ice: 2.5 µl RNA-sample (100–150 ng total RNA) + 1 µl 5× MMLV buffer + 1 µl RT primer mix. Heat for 1 min at 80°C, then incubate for 5 min at 45°C.
2. During the 5-min incubation, the MMLV reverse transcriptase enzyme is diluted from 200 to 20 U/µl using a 1:1 mix of water and SALSA Enzyme Dilution buffer (blue cap). This should be stored on ice prior to use.
3. Lower the temperature of the thermocycler to 37°C. Add 1.5 µl 20 U/µl MMLV reverse transcriptase and incubate for 15 min. Inactivate the reverse transcriptase by heating at 98°C for 2 min, then cool to 25°C.
4. Continue from step 3 in Subheading 3.1.

3.3. Methylation Specific MLPA (MS-MLPA)

1. Follow the standard MLPA protocol from steps 1–3 in Subheading 3.1.
2. After hybridization, dilute with H_2O and 3 µl Ligase buffer A to a final volume of 20 µl and then divide equally between two tubes at room temperature. Incubate both samples at 49°C for at least 1 min.
3. While at 49°C, add a mixture of 0.25 µl Ligase-65, 0.5 µl HhaI enzyme, 1.5 µl Ligase buffer B, and 7.75 µl H_2O to

one tube. A similar mixture is added to the second tube, with the only difference being H_2O instead of HhaI.

4. Perform ligation and digestion by incubating for 30 min at 49°C. Heat to 98°C for 5 min to inactivate the enzymes.
5. Continue from step 5 in Subheading 3.1.

3.4. Extension-MLPA

1. The protocol is very similar to standard MLPA described in Subheading 3.1. The only difference is at step 4 in Subheading 3.1. For each reaction, mix 3 µl Ligase-65 buffer A (transparent cap) + 3 µl Ligase-65 buffer B (white cap) + 25 µl H_2O + 0.2 µl dNTPs + 0.1 µl Stoffel fragment. Add 1 µl Ligase-65 (green cap) and mix again. This is added to the hybridization reaction while the PCR tubes are in the thermal cycler at 54°C.

3.5. DNaseI-MLPA

1. Harvest cells (~2×10^5–10^6) and isolate nuclei using the NE-PER Nuclear and Cytoplasmic Extraction kit. All solutions are provided with the kit (see Note 9).
2. Spin down cells at 500 × *g* for 2 min and discard the supernatant, leaving cell pellet as dry as possible.
3. Resuspend cell pellet in ice-cold CER I (100 µl for 10^6 cells, see Note 10) by flicking the tube or gently pipetting up and down. Incubate on ice for 10 min.
4. Add 5.5 µl of ice-cold CER II to the tube, mix by flicking the tube, and incubate on ice for 1 min.
5. Flick tube again and centrifuge at 250 × *g* at 4°C for 2 min and discard the supernatant (see Note 11).
6. Wash nuclei with 500 µl of ice-cold DNaseI reaction buffer (provided with the enzyme, containing 2% glycerol), centrifuge at 250 × *g* for 2 min at 4°C, and gently resuspend in 75 µl DNaseI reaction buffer containing 2% glycerol (at RT, see Note 12).
7. Add 25 µl nuclei aliquots to 50 µl of DNaseI solution (reaction buffer containing 2% glycerol and required amount of DNaseI) in a 2-ml Eppendorf tube, mix, and incubate for 20 min at RT.
8. Stop the reaction and release digested genomic DNA by adding 250 µl nuclei lysis buffer (containing 50 µg Proteinase K) and incubate at 55°C for 45 min.
9. Transfer tubes to 37°C; add 20 µg RNaseA and incubate at 37°C for 30 min.
10. Use the HighPure PCR purification kit to clean up and recover digested DNA (see Note 13).
11. Add 1.6 ml Binding Buffer and mix.

12. Transfer up to 700 µl to a High Pure filter tube and spin at maximum speed for 30 s. Discard flow-through and repeat step with remaining solution.
13. Add 500 µl wash buffer and centrifuge at maximum speed for 1 min, discard flow-through.
14. Add 200 µl wash buffer and centrifuge at maximum speed for 1 min, transfer filter tube to a clean 1.5-ml Eppendorf tube.
15. Add 50 µl elution buffer and centrifuge at maximum speed for 1 min.
16. Continue from step 1 in Subheading 3.1 (see Note 14).

4. Notes

1. A critical requirement for a successful MLPA reaction is good-quality template DNA. Although a degree of degradation can be tolerated, as the target region for most probes is <100 bp, impurities such as organic solvents can lead to inaccurate results. For best results, only DNA samples that were isolated using the same method should be normalized together.
2. Incomplete denaturation of the genomic template can mean that some probes do not have full access to the target sequences. It may be necessary to increase the length of denaturation to improve probe accessibility.
3. Use of plugged tips is recommended throughout the procedure.
4. Although the protocol uses an overnight (~16 h) hybridization time to allow the probes to anneal to the genomic template, there have been reports showing that 2–3 h is sufficient (14, 15). As probes may differ in annealing efficiency, shorter hybridization times should be tested for each probe mix before being routinely implemented.
5. The ligated products are stable and can be stored for months at –20°C.
6. To save on reagents, it is possible to reduce the volume of the PCR reaction.
7. This assumes the use of a capillary sequencer such as the ABI3700 (Applied Biosystems). Other machines may have different requirements.
8. The optimal amount of PCR product may be different, depending on the sensitivity of the specific sequencer.
9. To minimize changes in chromatin structure, nuclei should be kept on ice and the procedure performed as quickly as possible.

10. Reduce amount of solutions if using fewer than 10^6 cells; however, do not use less than 50 µl of CER I. Depending on the cell line, isolated nuclei can be sticky and very hard to resuspend. In this case, double amounts of NE-PER solution can be used. If unsuccessful, harvested cells can be aliquoted before nuclei isolation.
11. The integrity of the nuclei can be checked at this point by observation through a light microscope.
12. It is recommended to use two different DNaseI concentrations (e.g., 0.5 and 2 U) and an undigested control. If more conditions are required, resuspend the nuclei in a larger volume of reaction buffer.
13. It is important to efficiently purify and recover very long as well as very short fragments of DNA. Phenol–Chloroform extraction is not suitable for this task, as traces of phenol are known to interfere with MLPA.
14. If extension MLPA probes are used, the extension MLPA protocol (Subheading 3.5) should be used.

Acknowledgments

We thank Dr Jan Schouten and Professor Johan den Dunnen for their helpful discussions. SW is funded by NHMRC Australia (grant numbers 491293 and 546478).

References

1. Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F., and Pals, G. (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* **30,** e57.
2. Harteveld, C.L., Voskamp, A., Phylipsen, M., Akkermans, N., den Dunnen, J.T., White, S.J., and Giordano, P.C. (2005) Nine unknown rearrangements in 16p13.3 and 11p15.4 causing alpha- and beta-thalassaemia characterised by high resolution multiplex ligation-dependent probe amplification. *J Med Genet* **42,** 922–31.
3. White, S.J., Vink, G.R., Kriek, M., Wuyts, W., Schouten, J., Bakker, B., Breuning, M.H., and den Dunnen, J.T. (2004) Two-color multiplex ligation-dependent probe amplification: detecting genomic rearrangements in hereditary multiple exostoses. *Hum Mutat* **24,** 86–92.
4. Spek, C.A., Verbon, A., Aberson, H., Pribble, J.P., McElgunn, C.J., Turner, T., Axtelle, T., Schouten, J., Van Der Poll, T., and Reitsma, P.H. (2003) Treatment with an anti-CD14 monoclonal antibody delays and inhibits lipopolysaccharide-induced gene expression in humans in vivo. *J Clin Immunol* **23,** 132–40.
5. Eldering, E., Spek, C.A., Aberson, H.L., Grummels, A., Derks, I.A., de Vos, A.F., McElgunn, C.J., and Schouten, J.P. (2003) Expression profiling via novel multiplex assay allows rapid assessment of gene regulation in defined signalling pathways. *Nucleic Acids Res* **31,** e153.
6. Nygren, A.O., Ameziane, N., Duarte, H.M., Vijzelaar, R.N., Waisfisz, Q., Hess, C.J., Schouten, J.P., and Errami, A. (2005) Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. *Nucleic Acids Res* **33,** e128.
7. Leonhard, W.N., Roelfsema, J.H., Lantinga-van Leeuwen, I.S., Breuning, M.H., and Peters, D.J.

(2008) Quantification of Cre-mediated recombination by a novel strategy reveals a stable extra-chromosomal deletion-circle in mice. *BMC Biotechnol* **8,** 18.

8. Ohnesorg, T., Eggers, S., Leonhard, W.N., Sinclair, A.H., and White, S.J. (2009) Rapid high-throughput analysis of DNaseI hypersensitive sites using a modified MLPA approach. *BMC Genomics* **10,** 412.
9. Lalic, T., Vossen, R.H., Coffa, J., Schouten, J.P., Guc-Scekic, M., Radivojevic, D., Djurisic, M., Breuning, M.H., White, S.J., and den Dunnen, J.T. (2005) Deletion and duplication screening in the DMD gene using MLPA. *Eur J Hum Genet* **13,** 1231–4.
10. Zeng, F., Ren, Z.R., Huang, S.Z., Kalf, M., Mommersteeg, M., Smit, M., White, S., Jin, C. L., Xu, M., Zhou, D.W., Yan, J.B., Chen, M.J., van Beuningen, R., Huang, S.Z., den Dunnen, J., Zeng, Y.T., and Wu, Y. (2008) Array-MLPA: comprehensive detection of deletions and duplications and its application to DMD patients. *Hum Mutat* **29,** 190–7.
11. Notini, A.J., Craig, J.M., and White, S.J. (2008) Copy number variation and mosaicism. *Cytogenet Genome Res* **123,** 270–7.
12. White, S.J., Vissers, L.E., Geurts van Kessel, A., de Menezes, R.X., Kalay, E., Lehesjoki, A.E., Giordano, P.C., van de Vosse, E., Breuning, M.H., Brunner, H.G., den Dunnen, J.T., and Veltman, J.A. (2007) Variation of CNV distribution in five different ethnic populations. *Cytogenet Genome Res* **118,** 19–30.
13. Groth, M., Szafranski, K., Taudien, S., Huse, K., Mueller, O., Rosenstiel, P., Nygren, A.O., Schreiber, S., Birkenmeier, G., and Platzer, M. (2008) High-resolution mapping of the 8p23.1 beta-defensin cluster reveals strictly concordant copy number variation of all genes. *Hum Mutat* **29,** 1247–54.
14. Aten, E., White, S.J., Kalf, M.E., Vossen, R.H., Thygesen, H.H., Ruivenkamp, C.A., Kriek, M., Breuning, M.H., and den Dunnen, J.T. (2008) Methods to detect CNVs in the human genome. *Cytogenet Genome Res* **123,** 313–21.
15. Bayley, J.P., Grimbergen, A.E., van Bunderen, P.A., van der Wielen, M., Kunst, H.P., Lenders, J.W., Jansen, J.C., Dullaart, R.P., Devilee, P., Corssmit, E.P., Vriends, A.H., Losekoot, M., and Weiss, M.M. (2009) The first Dutch SDHB founder deletion in paraganglioma-pheochromocytoma patients. *BMC Med Genet* **10,** 34.

Chapter 14

Assessing Gene-Specific Methylation Using HRM-Based Analysis

Ee Ming Wong and Alexander Dobrovic

Abstract

As DNA methylation analysis enters the mainstream of biomedical research, there is increasing interest in methodologies that can be used for rapid analysis of a wide variety of biological and clinical samples. The methods that have been commonly used for methylation analysis are usually time-consuming and require multiple steps. The methylation-sensitive high-resolution melting (MS-HRM) assay is an in-tube assay which was developed to assess promoter methylation in sodium bisulphite-modified DNA and is particularly useful for assessing methylation in short fragments of DNA derived from formalin-fixed paraffin-embedded biopsies.

Key words: Methylation, CpG dinucleotide, Methylation-sensitive high-resolution melting analysis

1. Introduction

Detectable levels of methylation, in particular at the promoter region CpG island, are often associated with gene silencing, thereby resulting in no mRNA or reduced mRNA levels (1, 2). Many efforts have been made in evaluating promoter region methylation in genes associated with cancer development (such as *BRCA1*) by comparing methylation levels in DNA isolated from cancerous and non-cancerous cells (3, 4). Whereas genomic sequencing (5) remains the gold standard, it and other PCR-based methods are often time-consuming multiple-step processes. In addition, they require the transfer of amplified DNA products leading to significant PCR contamination issues.

Several laboratories have developed in-tube assays to simplify the detection of methylation in regions of interest. The MethyLight real-time PCR assay is a modified methylation-specific PCR assay

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_14, © Springer Science+Business Media, LLC 2011

(6) which uses a TaqMan® type probe to monitor amplification in real time (7, 8). The cycle threshold (C_t) values obtained are normalised relative to a control assay which amplifies a bisulphite-modified template that contains no CpG sites. The melting curve assay amplifies sequences regardless of their methylation status in the presence of the intercalating fluorescent dye, SYBR Green I (9). A subsequent melting analysis reveals the presence of methylated or unmethylated DNA between the primers based on the increased melting temperature when cytosine residues (corresponding to methylated cytosine residues) are present compared to when thymine residues (corresponding to unmethylated cytosine residues) are present. The HeavyMethyl assay is designed to detect the presence of methylation in limited quantities of DNA (picogram amounts) and utilises an oligonucleotide blocker, which anneals specifically to unmethylated DNA in the region of interest (10). In the presence of unmethylated DNA, the blocker prevents the binding of methylation-specific primers thereby avoiding the amplification of unmethylated DNA and false-positives. In the presence of methylated DNA, the blocker fails to bind thus allowing methylation-specific primers to anneal to and amplify methylated DNA. Methylation levels are then detected using a fluorescent probe.

The methylation-sensitive high-resolution melting (MS-HRM) assay was first described in 2007 (11). This method is based on high-resolution melting (HRM) analysis, a method which utilises an intercalating fluorescent dye that is capable of associating with double-stranded DNA in a saturating manner (12). The melting is performed on real-time PCR instrumentation capable of monitoring the melting at high resolution and using software designed for HRM analysis. During the melting step, the double-stranded DNA is denatured to single-stranded DNA and the intercalating dye disassociates from the single-stranded DNA and stops fluorescing. This decrease in fluorescence is measured and recorded as a melting curve.

Using MS-HRM, methylation of the region of interest can be determined by comparing the melting curve profiles of the DNA sample of interest against a standard curve. Moreover, MS-HRM allows the flexibility of detecting different methylation levels in DNA from different sources using the same primer pair. MS-HRM is also useful as it allows the direct visualisation of heterogeneous methylation. As individual alleles can hybridise with other alleles that differ at only one, two, or three positions, multiple heteroduplexes can be formed. Because of the complexity of these interactions, estimation of the amount of methylated alleles is difficult and is best performed digitally (13, 14). Methylation can be assessed using either normalised melting curves, difference curves or second derivative (Tm) curves. In our laboratory, we find that Tm curves are the most convenient way of looking at the analysed data.

2. Materials

2.1. Sodium Bisulphite Modification of DNA

1. Sodium bisulphite modification kit (see Note 1).
2. Fully methylated DNA (see Note 2).
3. Unmethylated DNA (see Note 3).

2.2. MS-HRM Analysis

1. Hotstart Taq DNA polymerase (5 U/μl) provided with 10× PCR Buffer and 25 mM $MgCl_2$ (Qiagen). Store at –20°C.
2. 2.5 mM dNTP mixture (see Subheading 3.4).
3. 5 μM Each of forward and reverse primers (see Subheading 3.4).
4. 4 μM Syto9 green-fluorescent nucleic acid stain (Invitrogen) (see Subheading 3.4). Syto9 should be stored at –20°C and in the dark.

3. Methods

3.1. Assay Design

3.1.1. Primer Design for Evaluation of Methylation Levels After Sodium Bisulphite Modification

1. Design primers to amplify a region of interest containing CpG dinucleotides which has been subjected to sodium bisulphite modification. Sequences can be modified artificially by replacing cytosine residues that are not part of a CpG dinucleotide with thymine residues in a text document (4) (Fig. 1).
2. MS-HRM works best if primers overlap at least one CpG dinucleotide (see Notes 4–6; Fig. 1). The inclusion of more than one CpG dinucleotide often cannot be avoided but may make the primers more specific for methylated sequences. Additionally, CpG dinucleotides should be located close to the 5′ end of primers while the 3′ end of primers should

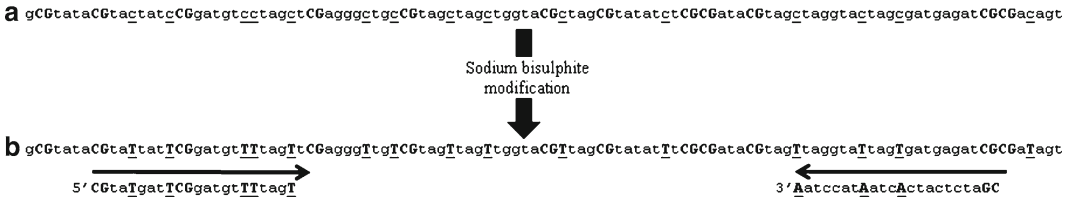


Fig. 1. Primer design for MS-HRM assay. Nucleotide sequence of a (hypothetical) CpG-rich gene region (**a**) before and (**b**) after bisulphite modification. CpG dinucleotides are in *bold caps*. During bisulphite modification, unmethylated cytosine residues (*underlined*) are converted to uracil residues and ultimately thymine residues (*bold caps* and *underlined*) during PCR. Methylated cytosine residues are not converted. Primers should anneal to at least one but no more than three CpG dinucleotides as close as possible to the 5′ end of the primer. Thymine residues (corresponding to unmethylated cytosine residues) should be positioned near the 3′ end of the primer to select against incompletely converted templates.

anneal to at least one thymine residue (corresponding to an unmethylated cytosine residue) (11, 15, 16).

3.1.2. Optimisation of Primer Annealing Temperature

1. The primer annealing temperature normally should be optimised to amplify both methylated and unmethylated sequences. We normally design our primers to have an annealing temperature of 65°C (salt-adjusted).
2. Run the initial MS-HRM assay at an annealing temperature of 60°C. To favour the amplification of methylated (cytosine-rich) sequences, the annealing temperature can be increased. Conversely, to favour the increased amplification of unmethylated (thymine-rich) sequences, the annealing temperature can be lowered (see Note 6; Fig. 2).

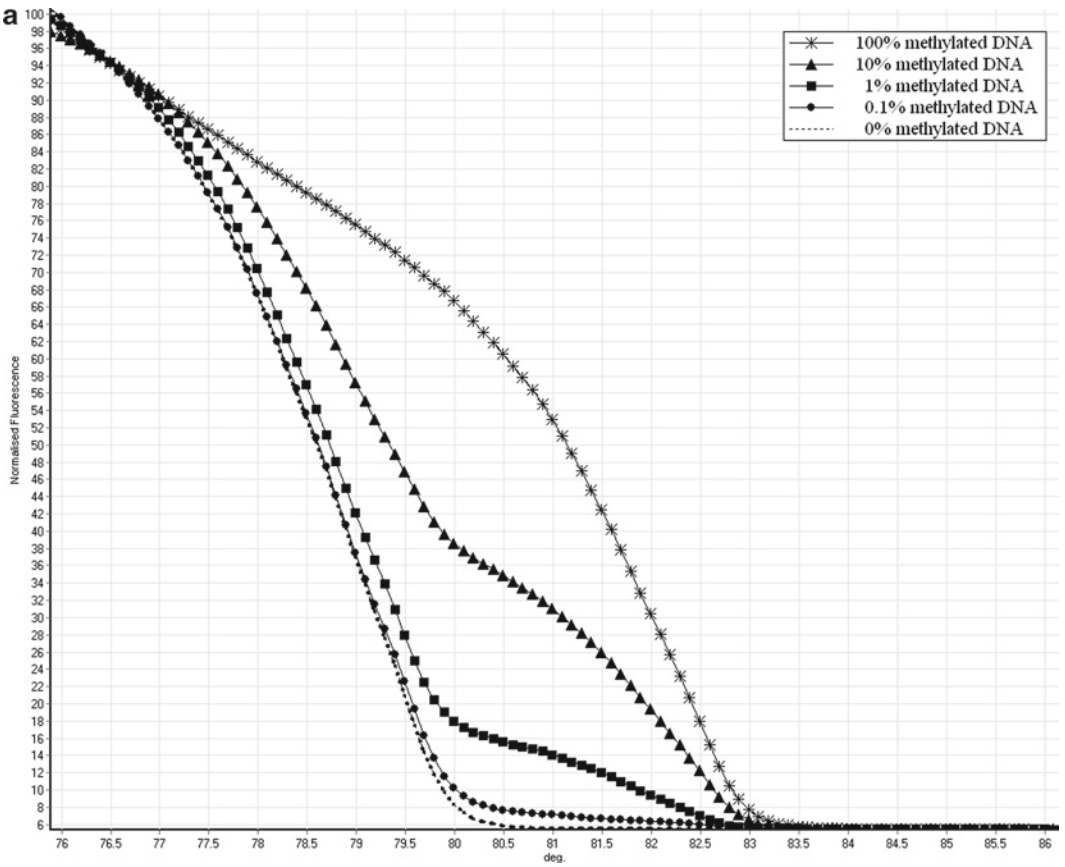
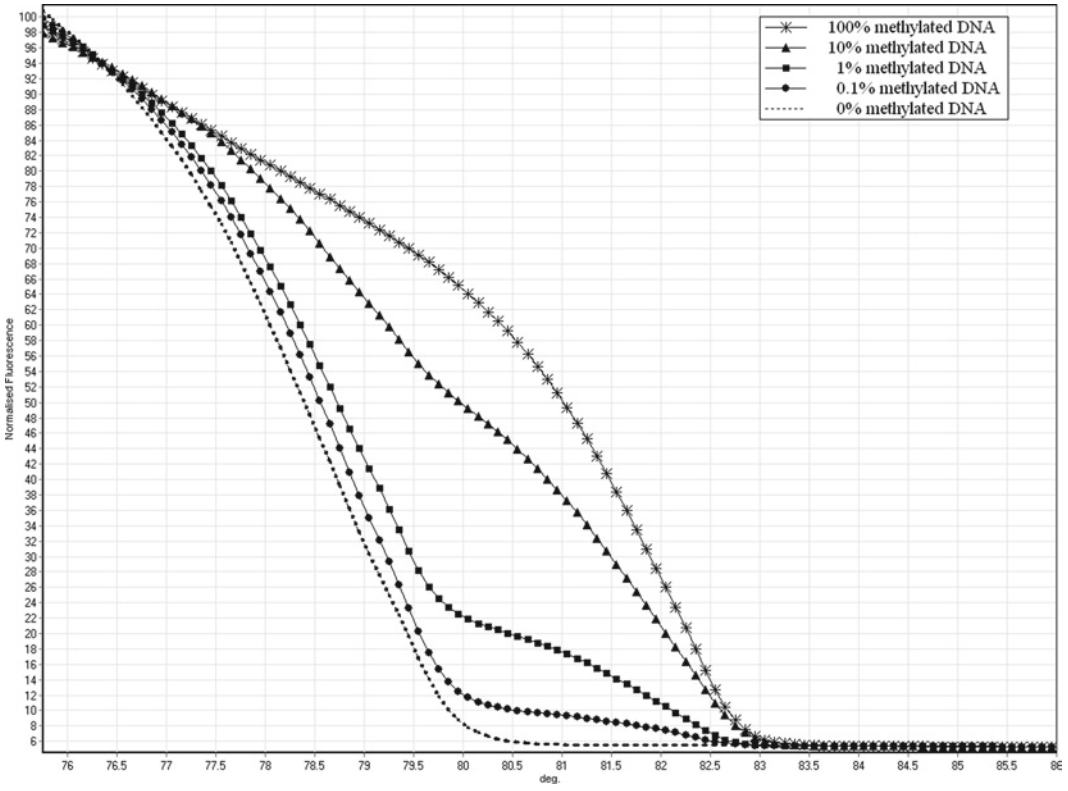


Fig. 2. Optimisation of primer annealing temperature. (**a**) PCR amplification at a primer annealing temperature of 60°C (*left panel*) resulted in a high-resolution melting profile which showed little discrimination between 0.1% methylated DNA (*circle*) and 0% methylated DNA (*dashed line*). By increasing the annealing temperature to 61°C (*right panel*) primers are *more* selective for cytosine-rich sequences (corresponding to methylated sequences) where increased differentiation between the 0.1 and 0% methylated DNA is observed. This is important where the detection of low-level methylation (0.1–5%) is desired. (**b**) PCR amplification at 60°C annealing temperature (*left panel*) resulted in a melting profile which showed little discrimination between 100% methylated DNA (*crosses*) and 10% methylated DNA (*triangle*). By lowering the annealing temperature to 59°C (*right panel*), primers were *less* selective for methylated sequences thus allowing for better separation between alleles containing higher levels of methylation. This is preferable where samples with moderate to high levels of methylation (5–100%) are being analysed.



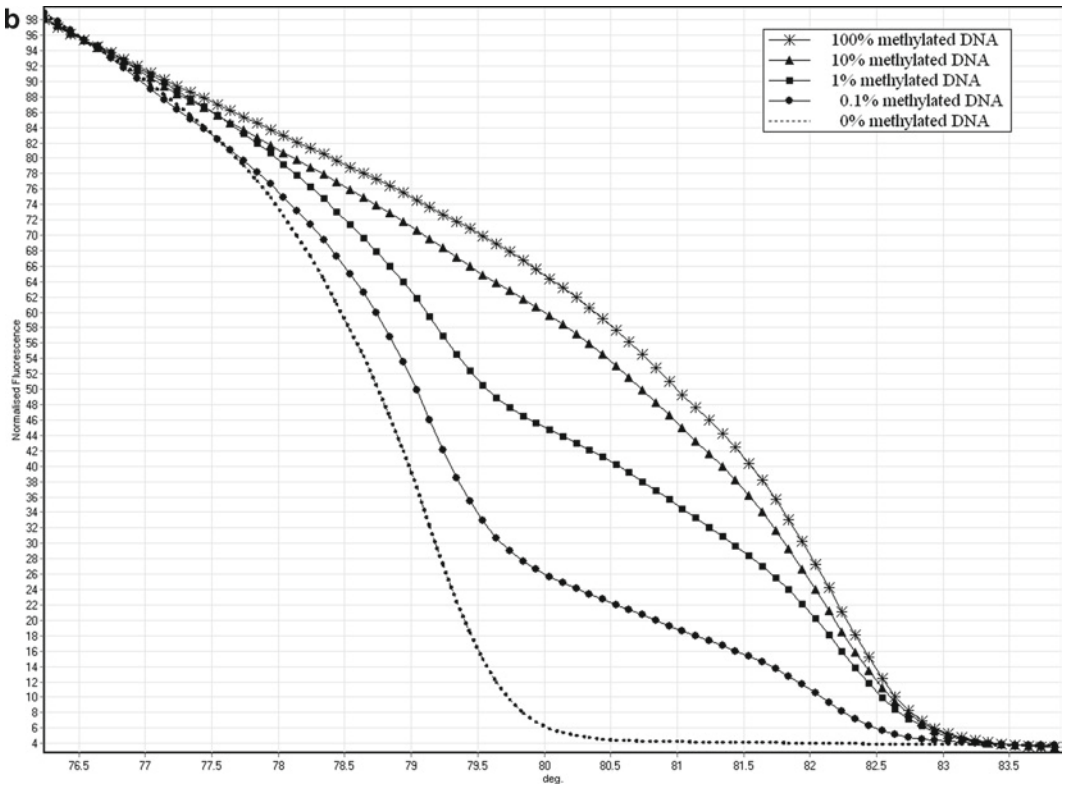


Fig. 2. (continued)

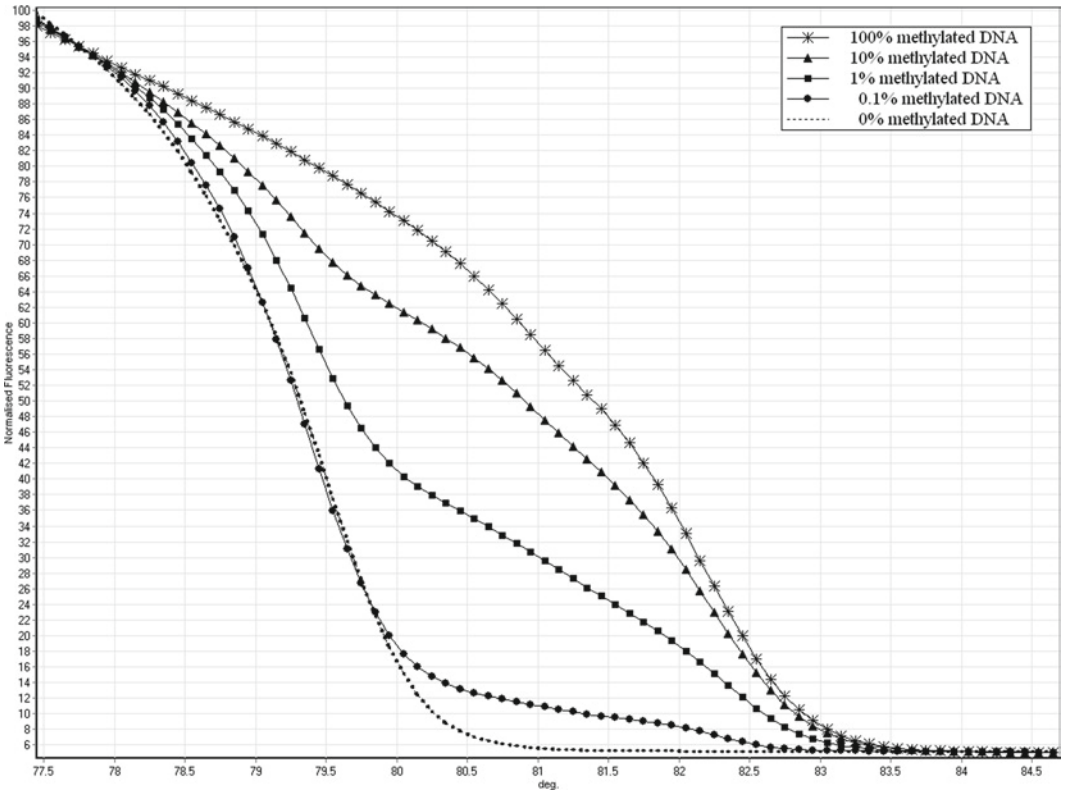


Fig. 2. (continued)

3.2. Sodium Bisulphite Modification of DNA Samples

1. Proceed with sodium bisulphite modification according to the manufacturer's protocol (see Note 7).
2. For each sodium bisulphite modification experiment, at least two aliquots of fully methylated DNA and a panel of unmethylated DNA should be bisulphite modified to reduce experimental variation.
3. The starting amount for all DNA samples should be the same to reduce experimental variation.
4. Resuspend bisulphite-modified DNA in elution buffer (provided by the manufacturer) or sterile water to a concentration of approximately 10 ng/μl (see Note 8).
5. Sodium bisulphite-modified samples can be stored long-term at −20 to −35°C without compromising DNA quality (see Notes 9 and 10).

3.3. Setting up a Standard Curve of "Percentage of Methylation"

1. After sodium bisulphite modification, pool aliquots of fully methylated DNA. Also pool aliquots of unmethylated DNA prior to setting up the relevant standard curves.
2. The standard curve of percentage of methylation is used to confirm the presence (and proportion) of DNA methylation

in the region of interest. Special care should be observed when setting up the standard curve. Dilute bisulphite-modified fully methylated DNA in bisulphite-modified unmethylated DNA to obtain approximate methylation percentages of (as an example): 100% methylated DNA (100 µl fully methylated DNA), 10% methylated DNA (10 µl 100% methylated DNA added to 90 µl unmethylated DNA), 1% methylated DNA (9 µl 10% methylated DNA added to 90 µl unmethylated DNA), 0.1% methylated DNA (9 µl 1% methylated DNA added to 90 µl unmethylated DNA), and 0% DNA (100 µl unmethylated DNA).

3.4. Assessment of Promoter Methylation Using MS-HRM

1. Prepare the reaction mix according to the following conditions: 1× PCR Buffer, 2.5 mM $MgCl_2$, 200 nM Syto 9, 500 nM each of forward and reverse primers, 200 µM dNTP mixture, 0.5 U Hotstart DNA Taq Polymerase, approximately 20 ng DNA, and sterile water up to a total volume of 20 µl (see Notes 11–13).
2. Perform MS-HRM using the following cycling conditions: activation of Taq polymerase at 95°C for 15 min followed by 50 cycles of denaturation at 95°C for 10 s, primer annealing at 60°C for 10 s, and primer elongation at 72°C for 20 s (see Note 14).
3. Acquire fluorescent data on the green channel (wavelength 470–510 nm). On the Rotorgene-Q, choose the "optimize acquiring function" which automatically determines the optimal gain for each assay.
4. For the first assay, use the default settings for high-resolution melting which is normally performed from 65 to 95°C, rising by 0.1°C increments and 1 s between increments. The HRM temperature range can be adjusted accordingly for subsequent assays, according to the melting profile of the particular assay.

4. Notes

1. We have used with success the MethylEasy™ and MethylEasy™ Xceed Rapid DNA Bisulphite Modification Kits (Human Genetic Signatures) and the Epitect Bisulphite Kit (Qiagen). Other laboratories report good results with EZ DNA Methylation™ Kits (Zymo Research).
2. We use CpGenome™ Universal Methylated DNA (Millipore).
3. An appropriate "unmethylated control" must be carefully chosen. To prevent possible experimental bias from using DNA extracted from only one control sample, we recommend using DNA extracted from a panel of controls

(6–8 different samples). This is particularly important as differences in the amount of low-level methylation between each control exists. As a further precaution we have also used whole genome amplified DNA to reduce the amount of methylation to below background levels. Alternatively, unmethylated DNA can also be purchased.

4. In general, short amplicons with single melting domains are preferable to long amplicons. When working with degraded material such as DNA from formalin-fixed paraffin-embedded tissues, it is preferable to use amplicons of less than 100 bp in length. These should frame a minimum of four CpG sites. Longer amplicons up to 200–250 bp may be used to assess whether novel genes are methylated but multiple melting domains are often present.
5. To validate the annealing temperature (salt-adjusted) and integrity of our primers, we use the web-based program Oligonucleotide Properties Calculator (http://www.basic.northwestern.edu/biotools/oligocalc.html). We use the Amplify 3 software program (http://engels.genetics.wisc.edu/amplify/) to confirm that the target sequence is amplified using our primers and to check for primer dimers. Other web-based primer design programs such as Primer3 (http://primer3.sourceforge.net/) may also be used.
6. We prefer to design primers according to the principles first laid out by Wojdacz and Hansen (15). These primers incorporate limited numbers of CpG hybridising residues at the 5′ end of the primers (Fig. 1). Whereas these primers differ markedly from MSP primers, they increasingly favour the amplification of methylated sequences as the annealing temperature is raised. This enables compensation for the PCR bias caused by the preferential amplification of unmethylated templates relative to methylated templates. Normally, the annealing temperature is varied until an equivalent amplification of methylated and unmethylated DNA is obtained. A second consequence of incorporating CpG dinucleotides in the primers is that a more methylation-specific-like assay can be obtained by further increasing the annealing temperature. This has the effect of increasing the sensitivity of the detection of methylation to as low as 1% or 0.1% (Fig. 2).
7. We have highlighted the main issues relating to bisulphite-modified DNA in this section. For additional notes on sodium bisulphite modification, refer to (17).
8. It is not possible to confidently determine the concentration of the DNA sample post-modification using a spectrophotometer. However, it is important to confirm that DNA is present post-modification. DNA quantitation via real-time

PCR can be achieved by (1) assessing the C_t value (corrected for by the amplification efficiency) or (2) designing primers to amplify a region free of CpG dinucleotides (preferably in a gene other than your gene of interest) and assaying the DNA of interest against a standard curve of known DNA amounts.

9. Some manufacturers recommend storing bisulphite-modified DNA at −80°C. However, we find that bisulphite-modified DNA is of better quality when stored at −20°C or −35°C. To minimise degradation caused by repeated freeze–thawing, an aliquot of the modified DNA sample stored at 4°C should be used as a working sample to preserve the quality of the remaining bisulphite-modified DNA sample.
10. MS-HRM reactions are ideally performed immediately after sodium bisulphite modification while the modified DNA samples are "fresh." These reaction mixes generally amplify in a more robust manner hence providing a more informative melting curve. Nevertheless, information can often be obtained from bisulphite-modified DNA samples that have been stored for an extensive period at −20 to −35°C.
11. To obtain a final concentration of 200 μM dNTPs, add 1.6 μl of 2.5 mM dNTP mixture to a total reaction volume of 20 μl. To obtain a final concentration of 500 nM primer, add 2 μl of 5 μM primer stock in a total reaction volume of 20 μl. To obtain a final concentration of 200 nM Syto9, add 1 μl of 4 μM Syto9 stock to a total reaction volume of 20 μl.
12. An increased concentration of sodium bisulphite-modified DNA is required in each reaction mix relative to what might be used in a normal PCR as bisulphite-modified DNA can be degraded. Use approximately 20 ng of DNA (based on the starting concentration after bisulphite modification) in each reaction.
13. After reaction set-up, it is possible to leave reaction mixes at 4°C overnight and in the dark without compromising the quality of the assay. MS-HRM can be performed on any machine that is capable of HRM analysis. We have conducted our analyses using the Rotorgene-Q (previously Rotorgene 6000) (Qiagen) and on the LightCycler® LC480 System (Roche).
14. Amplified DNA fragments can be run on an agarose gel to confirm that the amplicon is the fragment of interest. The presence of the intercalating dye will not interfere with the visualisation of the DNA fragment on the gel. Additionally, amplified DNA fragments can be directly sequenced after standard clean-up approaches to remove unused primers and nucleotides. We use Exo-SAP IT (USB Corporation) for PCR product clean-up.

Acknowledgements

EMW was supported by the Dora Lush Biomedical Postgraduate Scholarship from the National Health and Medical Research Council of Australia. Development and continued studies of MS-HRM were made possible by grants to AD from the National Health and Medical Research Council of Australia, the US Department of Defense Breast Cancer Research Program under award number W81XWH-06-1-0670, the Cancer Council of Victoria and the National Breast Cancer Foundation of Australia. Views and opinions of, and endorsements by the authors do not reflect those of the US Army or the Department of Defense.

References

1. Rice, J.C., Massey-Brown, K.S., and Futscher, B.W. (1998) Aberrant methylation of the BRCA1 CpG island promoter is associated with decreased BRCA1 mRNA in sporadic breast cancer cells. *Oncogene* **17**, 1807–12.
2. Herman, J.G., Latif, F., Weng, Y., Lerman, M.I., Zbar, B., Liu, S., Samid, D., Duan, D.S., Gnarra, J.R., Linehan, W.M., et al. (1994) Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A* **91**, 9700–4.
3. Kristensen, L.S. and Hansen, L.L. (2009) PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. *Clin Chem* **55**, 1471–83.
4. Dobrovic, A. (2005) Methods for Analysis of DNA Methylation, in *Molecular Diagnostics for the Clinical Laboratorian* (Coleman, W. B., and Tsongalis, G. J., Eds.) 2 ed., Humana Press, Totowa, NJ.
5. Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* **89**, 1827–31.
6. Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D., and Baylin, S.B. (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A* **93**, 9821–6.
7. Eads, C.A., Danenberg, K.D., Kawakami, K., Saltz, L.B., Blake, C., Shibata, D., Danenberg, P.V., and Laird, P.W. (2000) MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res* **28**, E32.
8. Trinh, B.N., Long, T.I., and Laird, P.W. (2001) DNA methylation analysis by MethyLight technology. *Methods* **25**, 456–62.
9. Worm, J., Aggerholm, A., and Guldberg, P. (2001) In-tube DNA methylation profiling by fluorescence melting curve analysis. *Clin Chem* **47**, 1183–9.
10. Cottrell, S.E., Distler, J., Goodman, N.S., Mooney, S.H., Kluth, A., Olek, A., Schwope, I., Tetzner, R., Ziebarth, H., and Berlin, K. (2004) A real-time PCR assay for DNA-methylation using methylation-specific blockers. *Nucleic Acids Res* **32**, e10.
11. Wojdacz, T.K. and Dobrovic, A. (2007) Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucleic Acids Res* **35**, e41.
12. Wittwer, C.T., Reed, G.H., Gundry, C.N., Vandersteen, J.G., and Pryor, R.J. (2003) High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem* **49**, 853–60.
13. Snell, C., Krypuy, M., Wong, E.M., Loughrey, M.B., and Dobrovic, A. (2008) BRCA1 promoter methylation in peripheral blood DNA of mutation negative familial breast cancer patients with a BRCA1 tumour phenotype. *Breast Cancer Res* **10**, R12.
14. Candiloro, I.L., Mikeska, T., Hokland, P., and Dobrovic, A. (2008) Rapid analysis of heterogeneously methylated DNA using digital methylation-sensitive high resolution melting: application to the CDKN2B (p15) gene. *Epigenetics Chromatin* **1**, 7.
15. Wojdacz, T.K. and Hansen, L.L. (2006) Reversal of PCR bias for improved sensitivity

of the DNA methylation melting curve assay. *Biotechniques* **41**, 274–8.

16. Wojdacz, T.K., Hansen, L.L., and Dobrovic, A. (2008) A new approach to primer design for the control of PCR bias in methylation studies. *BMC Res Notes* **1**, 54.
17. Warnecke, P.M., Stirzaker, C., Song, J., Grunau, C., Melki, J.R., and Clark, S.J. (2002) Identification and resolution of artifacts in bisulfite sequencing. *Methods* **27**, 101–7.

Chapter 15

Alu PCR

Maurizio Cardelli

Abstract

Alu PCR is a rapid and easy-to-perform "DNA fingerprinting" technique based on the simultaneous analysis of many genomic loci flanked by *Alu* repetitive elements, which allows the detection of genetic polymorphisms and mutations in human and primate genomes. In the protocol described in the present chapter, two fluorochrome-labelled primers complementary to *Alu* sequences are used to perform the PCR, and the amplification products are then analysed by capillary electrophoresis. The resulting complex electrophoretic pattern may show sample-to-sample variability due to insertion, deletion, or sequence change of *Alu* retrotransposons, or caused by length variation of the sequence interposed between two *Alus*.

Key words: *Alu* PCR, Inter-*Alu* PCR, *Alu*, Retrotransposable elements, Genomic fingerprinting, Genetic polymorphisms, Germline mutations, Somatic mutations

1. Introduction

Alu elements are non-autonomous retrotransposons present in high number of copies in primate genomes. In the human genome, these repetitive sequences are interspersed on all chromosomes with more than one million copies. A great variability in *Alu* density is observed across human genomic regions (1, 2), with a distribution showing two peaks at 0.1 *Alu*/kb and 1 *Alu*/kb (3). Thanks to their genome-wide distribution, *Alu* elements can be used as target loci to perform a PCR-based genome fingerprinting, known as *Alu* PCR or inter-*Alu* PCR. The method is characterised by an high information level compared to other DNA fingerprint techniques (4). *Alu* PCR was originally aimed to isolate human-specific DNA from human–mouse cell hybrids (5), but it has been successively used to detect somatic mutations, mainly in tumoral samples (6–9), germline genetic variability

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_15, © Springer Science+Business Media, LLC 2011

(10), and polymorphic loci involved in quantitative, multigenic traits (2, 11). In its earliest conception (5, 12), *Alu* PCR is a polymerase chain reaction conducted using a single, *Alu*-specific primer, followed by polyacrylamide gel electrophoresis of PCR products. A genomic locus can be PCR-amplified using an *Alu*-specific primer when it is flanked by two *Alu* elements with opposite orientation and separated by no more than a few kilobases (Fig. 1a). Consequently, a PCR performed on template genomic DNA using an *Alu*-specific primer produces a multitude of anonymous DNA fragments that can be revealed by electrophoretic separation. *Alu* PCR patterns typically show interindividual variability, due to different types of genetic polymorphisms: length variation of intervening sequences, de novo insertion of flanking *Alu* elements, deletions, translocations, and mutation of priming sites (4, 13, 14). Alu sequences are subdivided into different subfamilies based on diagnostic mutations and, inside each Alu subfamily, single elements can slightly differ from the subfamily consensus. Consequently, a single primer designed to target the Alu general consensus sequence is expected to anneal only to some Alu elements. For this reason, the simultaneous use of two or more *Alu*-specific primers in the PCR, which is proposed in the protocol described in the present chapter, increases the

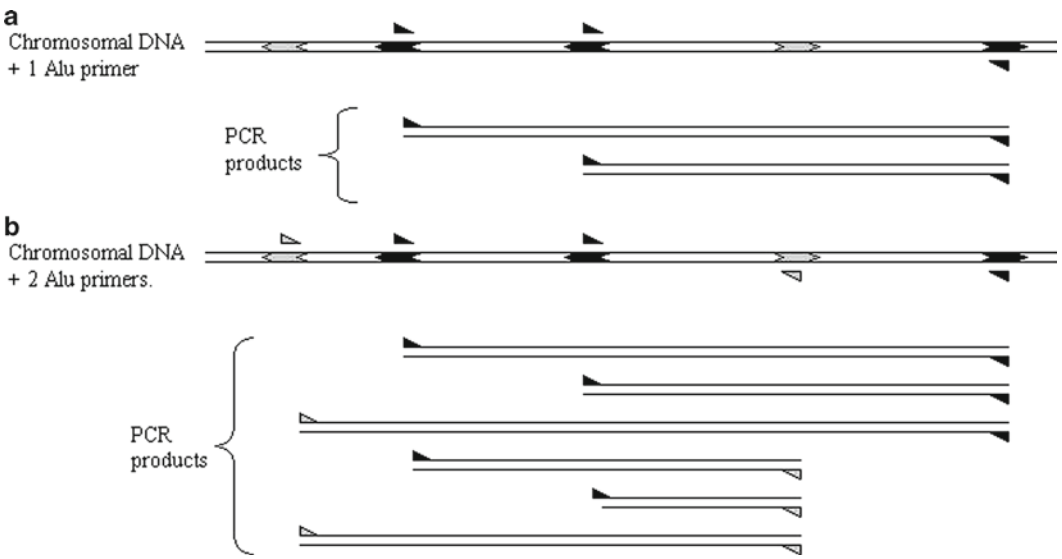


Fig. 1. Principle of *Alu* PCR. In (**a**) and (**b**), a chromosomal DNA region is represented, in which *Alu* sequences (*arrowed boxes*) are interspersed. *Alu* elements can be oriented in two possible orientations with respect to chromosomal DNA. (**a**) Because of sequence variation of *Alu* subfamilies and single *Alu* elements, when a primer (*black triangles*) designed to target *Alu* sequences is used in a PCR reaction, it is expected to anneal only to a subset of *Alu* sequences (represented in *black* on the chromosome). An amplification product is expected to be generated each time two primer-complementary *Alu* elements are in the correct reciprocal orientation (tail to tail, in the represented diagram) and no more distant than few kilobases. (**b**) When also a second *Alu*-targeted primer (*grey triangles*) is used in the PCR reaction, new annealing sites are found on the genome (*Alu* elements represented in *grey*), and the expected PCR products include fragments originated by two identical primer sites as well as fragments originated by two different primer sites.

number of loci that can be amplified by this technique (Fig. 1b) and produces a pattern of peaks which is more complex and information-rich with respect to that obtained by single-primer *Alu* PCR (see Note 1 and Fig. 1). The products resulting from *Alu* PCR are enriched in genomic sites located in *Alu*-rich regions, and *Alu* PCR could represent a genomic fingerprinting methodology focused on these peculiar genomic regions (2). Based on the same principle of *Alu* PCR, similar genomic fingerprinting methods can be designed targeting different families of high frequency interspersed repetitive sequences in the genomes of most eukaryote species.

2. Materials

2.1. Genomic DNA Extraction

1. QiaAmp DNA Blood (Qiagen).
2. Nanodrop ND 1000 spectrophotometer (Thermo Scientific).

2.2. Alu PCR

1. Oligonucleotide primers (see Note 1): R12A/267 (5′-AGCGAGACTCCG-3′), 5′ labelled with TET fluorochrome; R14B/264 (5′-CAGAGCGAGACTCT-3′) 5′ labelled with FAM fluorochrome. Primers are dissolved in nuclease-free distilled water at 10 μM, and stored at –20°C.
2. A dNTP mix is prepared containing 10 mM of each deoxynucleotide triphosphate (dATP, dCTP, dGTP, dTTP). It should be stored at –20°C.
3. Expand Tm High Fidelity PCR System DNA polymerase mix (Roche), and 10× Expand Tm High Fidelity Buffer (Roche), (12 mM $MgCl_2$, 20 mM Tris–HCl pH 7.5, 100 mM KCl, 1 mM dithiothreitol, 0.1 mM EDTA, 0.5% Tween 20 v/v, 0.5% Nonidet P40, 50% glycerol).
4. PCR is conducted using a 9700 thermal cycler (Applied Biosystems).

2.3. Capillary Electrophoresis

1. The electrophoresis is run on a 310 Genetic Analyzer (Applied Biosystems).
2. The “10× Genetic Analyzer Buffer with EDTA” (Applied Biosystems) is diluted 1:10 with distilled water before use, and used to fill the buffer vials of the genetic analyzer.
3. Pop 4 denaturing electrophoresis polymer and a 47 cm capillary (Applied Biosystems) are used for the electrophoretic runs (see Note 2).
4. GeneScan 2500, supplied by Applied Biosystems, is used as internal size standard.
5. Formamide ultrapure grade (Amresco, Solon, OH, USA) is used for sample preparation.

2.4. Data Analysis

1. The GeneScan program (supplied by Applied Biosystems) is used to visualize and to export the electrophoretic patterns.

3. Methods

This chapter describes an *Alu* PCR method based on the use of two fluorochrome-labelled primers, combined with capillary electrophoresis analysis of PCR products. The method has been presented in a previous paper (11) and used to detect a new longevity-associated locus (2). This variant of the *Alu* PCR technique uses two different *Alu*-specific primers labelled with different fluorochromes in the same PCR reaction; the resulting PCR products are then separated by capillary electrophoresis and fluorescent detection.

3.1. Genomic DNA Template

Human DNA samples are extracted from blood or other tissues, following standard procedures (see Note 3). QiaAmp DNA Blood is used in our laboratory for DNA extraction following the manufacturer instructions. After genomic DNA extraction/purification, it is suggested to check its quality using agarose gel electrophoresis, and to check DNA concentration using the Nanodrop ND 1000 spectrophotometer (see Note 4).

3.2. Alu PCR

Alu PCR is conducted using two primers, R12A/267 labelled with the "Tet" fluorochrome, and R14B/264 labelled with the "6-Fam" fluorochrome. PCR is carried out in a total volume of 50 µl with 20–200 ng of genomic DNA (see Note 5), 1× Expand Tm High Fidelity PCR buffer, 100 µM each deoxynucleotide triphosphate, 0.8 µM R12A/267 primer, 1.2 µM R14B/264 primer, 2.1 units of Expand Tm High Fidelity polymerase mix (see Note 6). PCR running conditions are the following: initial denaturation 5 min 94°C; 27 cycles at 94°C 30 s denaturation, 53.5°C 45 s annealing, 72°C 2 min extension; 7 min 72°C final extension. Samples should be stored at –20°C if not immediately analysed.

3.3. Capillary Electrophoresis

Alu PCR products are separated by capillary electrophoresis in a 310 Genetic Analyzer. Runs are performed in denaturing conditions, using Pop 4 electrophoresis polymer and 1× "GeneScan Buffer with EDTA" electrophoresis buffer.

1. Load 2 µl (see Note 7) of each *Alu* PCR amplified sample in a vial (0.5 ml vials should be used if the 310 Genetic Analyzer is provided with the 48 tube sample tray) containing 12 µl of formamide and 0.8 µl of GeneScan 2500 internal size standard.

2. The samples in formamide are then denatured at 95°C for 2 min in a thermal block, and then rapidly cooled on ice (keep samples on ice for at least 1 min).
3. Samples are loaded on the sample tray on the 310 Genetic Analyzer.
4. The sample injection is done at 15 kV for 15 s, while the electrophoretic run is conducted at 15 kV for 40 min.

3.4. Peak Analysis

The GeneScan program is used to visualize and to compare *Alu* PCR electrophoretic patterns. The *Alu* PCR patterns are very complex, and typically more than 100 peaks are resolved when the suggested combination of primers is used to perform the PCR reaction. An example of a complete *Alu* PCR pattern obtained with the combined use of R14B/264-Fam or R12A/267-Tet primers is shown in Fig. 2. The GeneScan program allows the analyst to zoom in on narrow regions of the electrophoretic pattern and to change the *y*-axis scale, to allow a visual analysis of smaller and/or more densely arranged peaks. The program also provides information about position, height, and area of each peak. In the GeneScan program, the colour of a peak indicates the labelling of the corresponding oligonucleotide. In the described protocol blue or green peaks correspond to single-strand oligonucleotides originated by the extension of R14B/264-Fam or R12A/267-Tet primers, respectively. When two peaks of identical length and different labelling are present, they are generally

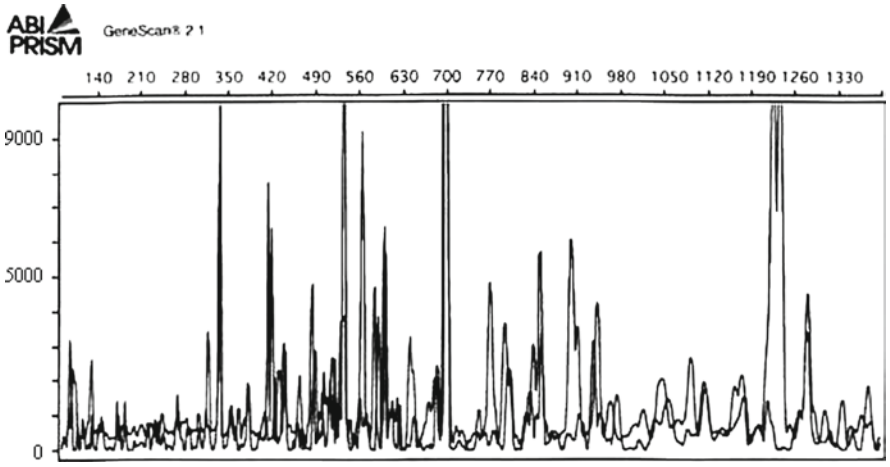


Fig. 2. The electrophoretic pattern of a double-primer *Alu* PCR analysed by capillary electrophoresis using the Applied Biosystems 310 automatic DNA sequencer: *y*-axis represents the relative fluorescence of the detected fragments, while the *x*-axis represents the fragment length (in base pairs). In this electropherogram the fragments span over a large size range (100–1,400 bp) and only higher peaks are evident. The GeneScan software allows the analyst to discern the peaks labelled by the two different fluorochromes (here not distinguished due to the *black* and *white* reproduction of the image).

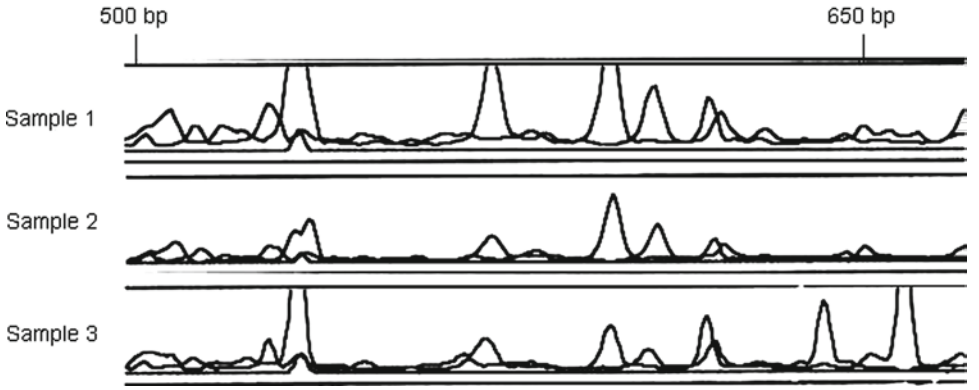


Fig. 3. A multi-alignment comparison of the electrophoretic pattern of three *Alu* PCR analysed samples in the region 500–670 bp is shown. The sample 3 shows two evident peaks (around 650 bp) that are not present in other samples.

due to an amplicon flanked by two different primers at the extremities (Fig. 1b). *Alu* PCR patterns of different samples can be compared by aligning different electropherograms (it can be done using the GeneScan program) to evidence peaks showing sample-to-sample variability (see Note 8). An example of a multi-alignment comparison of a region of the *Alu* PCR pattern is shown in Fig. 3.

4. Notes

1. Choice of the primers. The oligonucleotide primers used in the suggested double-primer *Alu* PCR protocol have been already used in some of the earliest single-primer *Alu* PCR protocols (4, 13). Both primers are nearly complementary, with forward orientation, to a region near (6–20 bp before) the 3′ end of the *Alu* consensus, with one (R14B/264) or two (R12A/267) different bases with respect to the consensus; R14B/264 is perfectly complementary to *Alu*Sz6 subfamily, while R12A/267 is perfectly complementary to the consensus of some of the major *Alu* subfamilies (such as *Alu*Sx and *Alu*Y subfamilies). It is important to note that the R12A/267 primer is also complementary to a region (about 380 bp from the 5′ end, with reverse orientation) of the consensus of human *L1* (*LINE-1*) retrotransposons, the other major family of human retrotransposons besides *Alu* sequences. As a consequence, the R12A/267 primer is potentially capable to generate not only *Alu–Alu*, but also *Alu-L1* and *L1–L1* amplicons, increasing the number and variety of polymorphisms and mutations that can be screened by this

method; however, due to the lower number of full-length *L1* elements respect to *Alu* elements, *Alu–Alu* amplicons are expected to be the prevalent products. The primers used in the present protocol are not designed to be human-specific: the same protocol and the same primers give specific *Alu* PCR patterns (different from human) when the genome of the New World monkey marmoset (*Callithrix jacchus*) is used as a template (unshown data); using other primate genomes, species-specific patterns can probably be obtained as well. It is important to note that different primers and primer combinations can be chosen, based on the specific aim of the experiment: for example, *Alu* PCR primers can be designed to preferentially target youngest (more polymorphic and mutagenic) subsets of *Alu* subfamilies, such as *Alu*Ya5 and *Alu*Ya8. The complexity of the profile is strictly dependent on the choice of primers: oligonucleotides designed to match the Alu consensus in a region of perfect identity between the various Alu subfamilies are expected to yield more numerous amplification products than primers designed to selectively target one or few subfamilies; primers designed to specifically target small subfamilies, such as Ya8 or Yb8, are expected to yield much fewer amplification products than primers targeting wide subfamilies, such as Sx. Consensus sequences of *Alu* subfamilies and of other repetitive elements in human and other genomes are publicly available at Repbase Update (15) website: http://www.girinst.org.

2. The capillary should be replaced every 100 runs.
3. DNA obtained by multiple displacement amplification (16) (MDA) of genomic DNA aliquots has been also tested in our laboratory as an alternative to "native" genomic DNA as a template for *Alu* PCR. MDA or other WGA (whole genome amplification) methods are useful when only small quantities of precious genomic DNA samples are available. Albeit the majority of electrophoretic peaks obtained with MDA samples are generally identical to those obtained by amplifying "native" genomic DNA, a small number of peaks obtained from "native" DNA are not observed when MDA DNA is used. On the whole, it seems that MDA DNA is a feasible template for *Alu* PCR, but samples obtained using "native" DNA and MDA DNA should not be directly compared.
4. Good quality genomic DNA runs as a major band of at least 10 kb on the gel, and has an A260/A280 ratio between 1.8 and 2.0 when read on the spectrophotometer.
5. The method is quite robust and not too sensitive to variations in template DNA quantity. The genomic DNA quantity which is here suggested (20–200 ng) seems to be the optimal range. In general, it has been observed that a relatively higher

template DNA quantity tends to favour the production of short PCR products at the expense of longer amplicons: in this case, in the electrophoretic pattern, the peaks corresponding to fragments below 1,000 bp tend to be higher and more evident than peaks corresponding to longer fragments. Using lower template DNA quantity (20–50 ng) tends to yield more even peak heights in a wide range of product lengths (30–1,700 bp).

6. Compared to "classical" singleplex PCR protocols, Alu PCR is more sensitive to variations in PCR conditions; it is advisable to avoid any change in PCR conditions, including type and brand of DNA polymerase and model of thermal cycler, during the execution of a set of experiments.
7. The volume (2 µl) of *Alu* PCR product that should be loaded in formamide is two to ten times higher than the loading volume which is necessary when typical singleplex PCR products are analysed on the same instrument.
8. The Alu PCR peaks are characterised by variable heights (depending on the different amounts of the various amplicons present at the end-point), but are clearly higher and generally easily distinguishable from the background noise. However, a minority of the peaks obtained in Alu PCR appear to be not reproducible, probably because they represent PCR products which are very sensitive to light and unpredictable variation in PCR conditions. Consequently, before beginning the analysis of experimental samples, we suggest to conduct a preliminary assay in a set of repeated samples, aimed to identify the unreproducible peaks: those peaks whose presence/absence is erratic among a series of repetitions of a same sample must be excluded from the analysis of all the experimental samples. Albeit the analysis and interpretation of an *Alu* PCR pattern is perfectly feasible by visual examination of electropherograms and associated data, the analysis of a wide number of samples is time-consuming. It can be useful to consider that such an analysis can be automated by implementing a computer algorithm based on wavelet analysis (Cardelli M. et al., manuscript in preparation).

Acknowledgements

The author would like to thank I.N.R.C.A. (Italian National Research Center on Aging), Polo Scientifico e Tecnologico, Ancona, for having supported the development of this analytical method.

References

1. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., et al. (2001) Initial sequencing and analysis of the human genome. *Nature*. **409**, 860–921.
2. Cardelli, M., Marchegiani, F., Cavallone, L., Olivieri, F., Giovagnetti, S., Mugianesi, E., et al. (2006) A polymorphism of the YTHDF2 gene (1p35) located in an *Alu*-rich genomic domain is associated with human longevity. *J Gerontol A Biol Sci Med Sci*. **61**, 547–56.
3. Moyzis, R.K., Torney, D.C., Meyne, J., Buckingham, J.M., Wu, J.-R., Burks, C., et al. (1989) The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics*. **4**, 273–89.
4. Jarnik, M., Tang, J.Q., Korab-Laskowska, M., Zietkiewicz, E., Cardinal, G., Gorska-Flipot, I., et al. (1996) Overall informativity, OI, in DNA polymorphisms revealed by inter-Alu PCR: detection of genomic rearrangements. *Genomics*. **36**, 388–98.
5. Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramírez-Solis, R., Webster, T.D., Ledbetter, D.H., and Caskey, C.T. (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc Natl Acad Sci U S A*. **86**, 6686–90.
6. Krajinovic, M., Richer, C., Labuda, D., and Sinnett, D. (1996) Detection of a mutator phenotype in cancer cells by inter-Alu polymerase chain reaction. *Cancer Res*. **56**, 2733–7.
7. Furmaga, W.B., Cole, S.R., and Tsongalis, G.J. (2004) The use of Alu PCR to distinguish between typical pulmonary carcinoids versus classic midgut carcinoids. *Int J Oncol*. **24**, 223–6.
8. Srivastava, T., Seth, A., Datta, K., Chosdol, K., Chattopadhyay, P., and Sinha, S. (2005) Inter-alu PCR detects high frequency of genetic alterations in glioma cells exposed to sublethal cisplatin. *Int J Cancer*. **117**, 683–9.
9. McKie, A.B., Iwamura, T., Leung, H.Y., Hollingsworth, M.A., and Lemoine, N.R. (1997) Alu-polymerase chain reaction genomic fingerprinting technique identifies multiple genetic loci associated with pancreatic tumourigenesis. *Genes Chromosomes Cancer*. **18**, 30–41.
10. Charlieu, J.P., Laurent, A.M., Carter, D.A., Bellis, M., and Roizès, G. (1992) 3′ Alu PCR: a simple and rapid method to isolate human polymorphic markers. *Nucleic Acids Res*. **20**, 1333–7.
11. Bonafè, M., Cardelli, M., Marchegiani, F., Cavallone, L., Giovagnetti, S., Olivieri, F., et al. (2001) Increase of homozygosity in centenarians revealed by a new inter-Alu PCR technique. *Exp Gerontol*. **36**, 1063–73.
12. Sinnett, D., Deragon, J.M., Simard, L.R., and Labuda, D. (1990) Alumorphs-human DNA polymorphisms detected by polymerase chain reaction using Alu-specific primers. *Genomics*. **7**, 331–4.
13. Zietkiewicz, E., Labuda, M., Sinnett, D., Glorieux, F.H., and Labuda, D. (1992) Linkage mapping by simultaneous screening of multiple polymorphic loci using Alu oligonucleotide-directed *PCR*. *Proc Natl Acad Sci U S A*. **89**, 8448–51.
14. Mighell, A.J., Markham, A.F., and Robinson, P.A. (1997) Alu sequences. *FEBS Lett*. **417**, 1–5.
15. Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005) Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res*. **110**, 462–7.
16. Blanco, L., Bernad, A., Lazaro, J.M., Martin, G., Garmendia, C., and Salas, M. (1989) Highly efficient DNA synthesis by the phage φ 29 DNA polymerase. Symmetrical mode of DNA replication. *J Biol Chem*. **264**, 8935–40.

Chapter 16

Asynchronous PCR

Caifu Chen, David Ruff, and Jason Halsey

Abstract

Asynchronous PCR (aPCR) is a new PCR method that directs an ordered and sequential amplification of the + and – strands of DNA amplicons. There are several unique characteristics of aPCR that generate new application opportunities. The melting temperature (Tm) of the forward and reverse aPCR primers differ by at least 15°C. The concentration of the lower Tm primer is reduced from 900 to 100 nM, thereby allowing for asynchronous or asymmetric strand-specific amplification. Furthermore, unique thermocycling parameter strategy dictates the + and – strand amplification cue. Each aPCR cycle includes two annealing and extension steps. Sequential annealing and extension of forward and reverse primers during each cycle produce transient single-stranded DNA (ssDNA) amplicons which help hybridization-based probes such as peptide nucleic acid (PNA) bind to the target sequences more effectively. This new method can be used in real-time quantitative PCR (qPCR) for gene expression analyses as well as production of robust ssDNA targets for microarray and other hybridization-based applications.

Key words: Asynchronous PCR, qPCR, Peptide nucleic acid, TaqMan® assays, Gene expression

1. Introduction

Real-time quantitative PCR (qPCR) is the technique of collecting amplification signals during each PCR cycle. The first attempt to monitor the PCR amplification process exploited the 5′-nuclease activity of *Taq* polymerase. Holland et al. in 1991 demonstrated that cleavage of a radioactively labeled probe during PCR by *Taq* DNA polymerase could be used to detect amplification of the target-specific product (1). In these experiments, an oligonucleotide probe that hybridized to the amplicon between the two PCR primers was labeled with ^{32}P on its 5′-end and a blocker at its 3′-end to prevent priming. During the amplification cycling, annealing of the probe to its target sequence generated a substrate that was cleaved by the 5′-nuclease activity as the polymerase extended from an upstream primer into the region of the

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_16, © Springer Science+Business Media, LLC 2011

probe. However, this monitoring required extensive manual sample manipulation at each cycle. The first documentation of real-time PCR amplification monitoring by fluorescence sensing was described by Higuchi et al. in 1992 (2). This real-time PCR approach externally measured the fluorescence of an individual PCR well. Addition of the dsDNA intercalating dye ethidium bromide allowed for a fluorescence signal production from the PCR amplification process. To date, there are three major classes of reporter technologies for qPCR which include intercalating dyes (3), fluorescence-labeled primers (4–6), and fluorescence-labeled probes (1, 7–10). The most widely used method is the 5′-nuclease assay or TaqMan® assay (7). Novel fluorogenic probes first developed by Lee et al. greatly hastened the development of TaqMan® qPCR chemistry for SNP genotyping and gene expression analysis (9, 11).

Peptide nucleic acid (PNA) is a DNA analog with high binding affinity and specificity (12). Recently, linear PNA beacon probes have been developed for real-time PCR (13, 14). PNA is a unique class of chimeric oligomer: a peptide backbone with protruding nucleic acid bases. PNA probes do not require a stem to facilitate quenching as the oligomer in its unbound state is highly folded. The possible advantages of employing linear beacons in real-time PCR analysis include the improved kinetics, higher inherent specificity, and the ability to use considerably shorter length probes than molecular beacons.

Even though other noncleavable probe moieties have been applied in real-time qPCR applications (15–17), such probes including the PNA probes usually do not give significant signals and robust amplification curves. We found that it was partly caused by low hybridization efficiency of these probes to double-strand DNA targets (Fig. 1). To address this issue, we employ a new asynchronous PCR (aPCR) strategy to preferentially amplify one of the DNA strands first during each cycle of PCR (see *US patent No.: US 6887664*). Furthermore, we demonstrate the broad utilities of aPCR for real-time quantification and generation of the end-point single-stranded PCR products.

2. Materials

2.1. Asynchronous PCR Primer and Fluorogenic Probe Sequences (Examples)

1. Synthetic human LIG1 target

   Standard qPCR primers and probes:

   Forward: CGATCCCGCTTGTGATACAGA (Tm = 62°C)

   Reverse: CCGGCTGGTCCCCGTC (Tm = 63°C)

   TaqMan® probe: (6-FAM)-CTGTTGCCACTTCAGCC-(MGB)

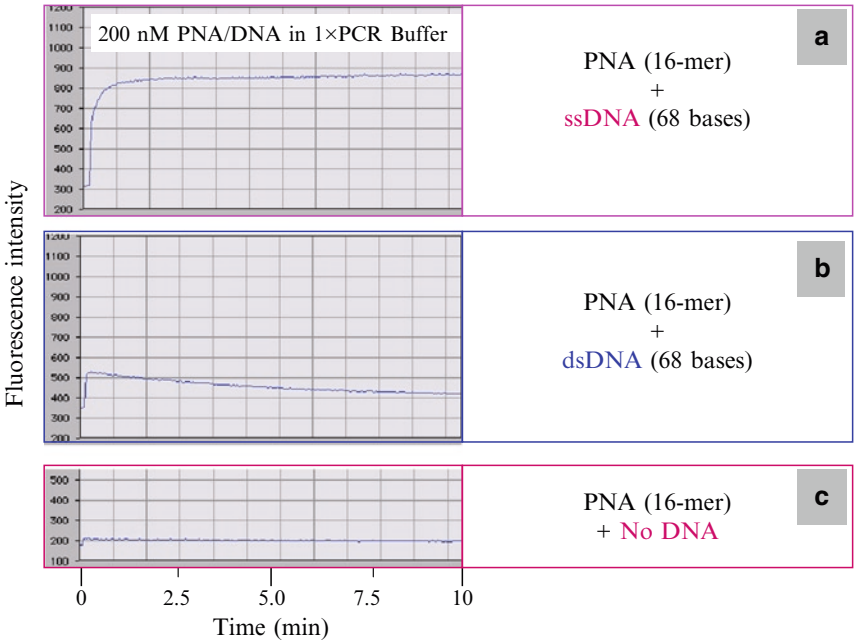


Fig. 1. Binding kinetics of PNA probes to single-stranded and double-stranded LIG1 targets. Double-labeled PNA oligomer was added to single-stranded (**a**), double-stranded (**b**) templates and no target DNA (buffer alone) (**c**), respectively. After denaturing at 95°C for 10 min and cooling down to 60°C, fluorescence signals were measured in a real-time qPCR instrument. An increase in fluorescence represented the binding event of PNA probe to its target over 10 min. There was no fluorescence change in the absence of target. PNA probe bound to single-stranded target very effectively, within a minute. However, in the presence of double-stranded target, PNA probe did not bind to the target effectively, resulting in much lower observed fluorescence signals and a decreased signal over time.

PNA: (Flu)-(Glu)-CTGTTGCCACTTCAGCC-(Lys-Lys-Dabcyl-NH_2)

aPCR primers and probes:

Forward: GCGCTGCGATCCCGCTTGTGATACAGA (Tm = 74°C)

Reverse: GGCTGGTCCCCGTC (Tm = 53°C)

TaqMan® probe: (6-FAM)-CTGTTGCCACTTCAGCC-(MGB)

PNA: (Flu)-(Glu)-CTGTTGCCACTTCAGCC-(Lys-Lys-Dabcyl-NH_2)

(see Note 1).

2. K-ras

Standard qPCR primers and probes:

Forward: GGATATCTGCAGAATTCGGCTTAT (Tm=62°C)

Reverse: CTCTATTGTTGGATCATATTCGTCCAC (61°C)

TaqMan®probe: (6-FAM)-ACGCCACCAGCTCCA-(MGB)

PNA: (Flu)-(Glu)-ACGCCACCAGCTCCA-(Lys-Lys-Dabcyl-NH_2)

aPCR primers and probes:

Forward: GGCGCCGGATATCTGCAGAATTCGGCTTAT (Tm = 74°C)

Reverse: GGATCATATTCGTCCACA (Tm = 53°C)

TaqMan® probe: (6-FAM)-ACGCCACCAGCTCCA-(MGB)

PNA: (Flu)-(Glu)-ACGCCACCAGCTCCA-(Lys-Lys-Dabcyl-NH_2)

3. Synthetic targets for end-point labeling

Standard PCR primers:

Forward: (Cy5 or Cy3)-TGCGATCCCGCTTGTGATAC (Tm = 62°C)

Reverse: GGCTGGTCCCCGTCTTCTCCT (Tm = 61°C)

aPCR primers:

Forward: (Cy5 or Cy3)-GCGCTGCGATCCCGCTTGTGA TACAGA (Tm = 74°C)

Reverse: GGCTGGTCCCCGTCT (Tm = 54°C)

2.2. Asynchronous PCR for PNA Assays

(*NOTE: Unless otherwise indicated, reagents and consumable materials are from Applied Biosystems, Foster City, CA*)

1. Genomic DNA or cDNA (see Note 2).
2. 20× PNA beacon assays (2 μM of forward and 18 μM of reverse PCR primers and 4 μM of forward linear PNA beacon probe).
3. TaqMan® 2× Universal PCR Master Mix, No AmpErase® UNG.
4. Nuclease-free water.
5. ABI PRISM® 384-well Clear Optical Reaction Plates.
6. ABI PRISM® Optical Adhesive Covers.
7. Applied Biosystems 7900HT Fast Real-Time PCR Instrument and User Guide.

2.3. Asynchronous PCR for TaqMan® Assays

1. Genomic DNA or cDNA (see Note 2).
2. 20× TaqMan® assays (2 μM of forward and 18 μM of reverse PCR primers and 5 μM of forward TaqMan® probe).
3. TaqMan® 2× Universal PCR Master Mix, No UNG.
4. Nuclease-free water.
5. ABI PRISM® 384-well Clear Optical Reaction Plates.
6. ABI PRISM® Optical Adhesive Covers.
7. Applied Biosystems 7900HT Fast Real-Time PCR Instrument and User Guide.

2.4. Generation of Single-Strand DNA Targets

1. Plasmid DNA or cDNA (see Note 2).
2. 20× PCR primer mix including 2 and 18 μM of forward and reverse PCR primers, respectively.
3. 2× AmpliTaq Gold® 360 Master Mix.
4. Nuclease-free water.
5. SYBR® Green II Gel Staining Solution (Molecular Probes, Eugene, OR; see Note 3).
6. 384-well PCR plates.
7. Applied Biosystems 9700 PCR Instrument.
8. 10% Novex TBU Denaturing Gels (Invitrogen, Carlsbad, CA).
9. Novex pre-cast gel system (Invitrogen, Carlsbad, CA).
10. AlphaImaging gel documentation system (Alpha Innotech Corp., San Leandro, CA).

2.5. Microarray Hybridization

1. Hybridization mixture (4× SSC, 0.3% SDS, 1 μg/μL of yeast tRNA, 1 μg/μL of poly(A) RNA, and 1–2 μL of 50-μL PCR product).
2. Microarray wash buffer I (4× SSC and 0.3% SDS).
3. Microarray wash buffer II (1× SSC and 0.3% SDS).
4. Microarray wash buffer III (0.06× SSC).
5. Microcon YM-100 (Millipore, Billerica, MA).
6. Axon slide scanner (Molecular Devices, Sunnyvale, CA).
7. GenePix Pro 3.0 software (Molecular Devices, Sunnyvale, CA).

3. Methods

3.1. Standard TaqMan® PCR

3.1.1. Assay Design

The Primer Express® Software Version 2.0 is used to design PCR primers and TaqMan® probes with MGB using the default assay design setting (see Note 4). Visual OMP™ Nucleic Acid Software (DNA Software, Ann Arbor, MI) is used to design PNA oligomers.

3.1.2. PCR Protocol

1. Dilute genomic DNA or cDNA to appropriate concentration (see Note 2).
2. Prepare PCR mix below:

PCR mix	μL
2× Universal TaqMan Master Mix	5
20× TaqMan Gene Expression Assay	0.5
dH_2O	0.5
Total	6

3. Add 6 μL of PCR mix above into each well in a 384-well plate.
4. Add 4 μL of the diluted DNA sample into each well in a 384-well plate.
5. Seal the plate using optical membrane, mix by inverting plates three times, and spin at 2,000 rpm for 30 s.
6. Run qPCR on the Applied Biosystems 7900HT Fast instrument according to the following thermocycling protocol (default):
   - 50°C for 2 min.
   - 95°C for 10 min.
   - 40 cycles of incubation at 95°C for 15 s and 60°C for 60 s.
7. Record the Ct values for each well based on the following steps:
   - View the amplification plots for the entire plate.
   - Set the baseline and threshold values (automatic baseline and threshold of 0.2 is recommended).
   - Export the data file and analyze using Microsoft Excel or other qPCR data analysis software.

3.2. Asynchronous PCR

3.2.1. Assay Design

The Primer Express Software Version 2.0 is used to design aPCR primers and MGB TaqMan® probes based on the following application parameter inputs (see Note 4). Visual OMP™ Nucleic Acid Software (DNA Software, Ann Arbor, MI) is used to design PNA oligomers.

	Tm (°C)	Concentration (μM)
Forward primer	74 (72–76)	0.9
Reverse primer	54 (52–56)	0.2
PNA probe	72 (70–75)	0.25
MGB TaqMan probe	70 (68–72)	0.25

3.2.2. aPCR Protocol (Fig. 2)

1. Dilute DNA samples to appropriate concentration (see Note 2).
2. Prepare PCR mix below:

PCR mix	μL
2× Universal TaqMan Master Mix	5
20× PNA Probe and Primer Mix	0.5
dH_2O	0.5
Total	6

3. Add 6 μL of PCR mix above into each well in a 384-well plate.
4. Add 4 μL of the diluted DNA sample into each well in a 384-well plate.
5. Seal the plate using optical membrane, mix by inverting plates three times, and spin at 1,000 ×g for 30 s.

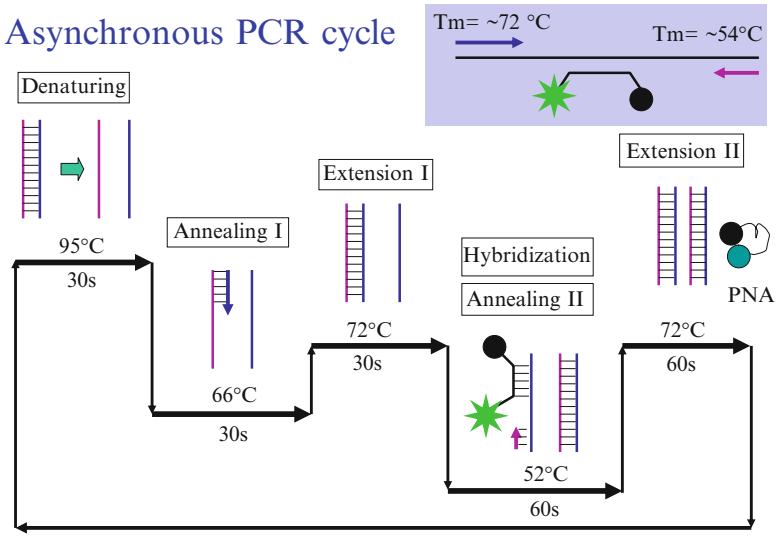


Fig. 2. Schematic of asynchronous PCR (aPCR) cycling protocol. aPCR included two annealing and denaturing steps: (1) denaturing double-stranded target, (2) annealing a first primer, (3) extension of the first primer, (4) probe hybridization, (5) annealing a second primer, and (6) extension of the second primer to make double-stranded amplicons. This unique cycling condition generates transient single-stranded DNA during each PCR cycle and thus allows PNA probe to bind to the target sequence effectively.

6. Run real-time qPCR on the Applied Biosystems 7900HT Fast instrument according to the following thermocycling protocol:
   - 95°C for 10 min.
   - 40 cycles of incubation at 95°C for 30 s, 66°C for 30 s (for forward priming), 72°C for 30 s, 52.5°C for 60 s (for reverse priming), and 72°C for 60 s.
7. Record the Ct values for each well based on the following steps:
   - View the amplification plots for the entire plate.
   - Set the baseline and threshold values (automatic baseline and auto Ct is recommended).
   - Export the data file and analyze using Microsoft Excel or other qPCR data analysis software.
   - Relative quantitation of a DNA target based on a standard curve method (Fig. 3.; also see Note 5).

3.3. Generation of Single-Strand DNA Targets for Microarrays

Single-strand DNA (ssDNA) targets could be generated by asymmetric PCR (18). ssDNA products (+ or – strands) can be effectively visualized in 10–15% denaturing polyacrylamide gel electrophoresis (PAGE, see Note 6) using SYBR® Green II Staining (see Note 3) and mobility modifier tagging in one of the PCR primers (see Note 7). However, the efficiency of asymmetric PCR is low largely due to its limited PCR primer. It often results in a

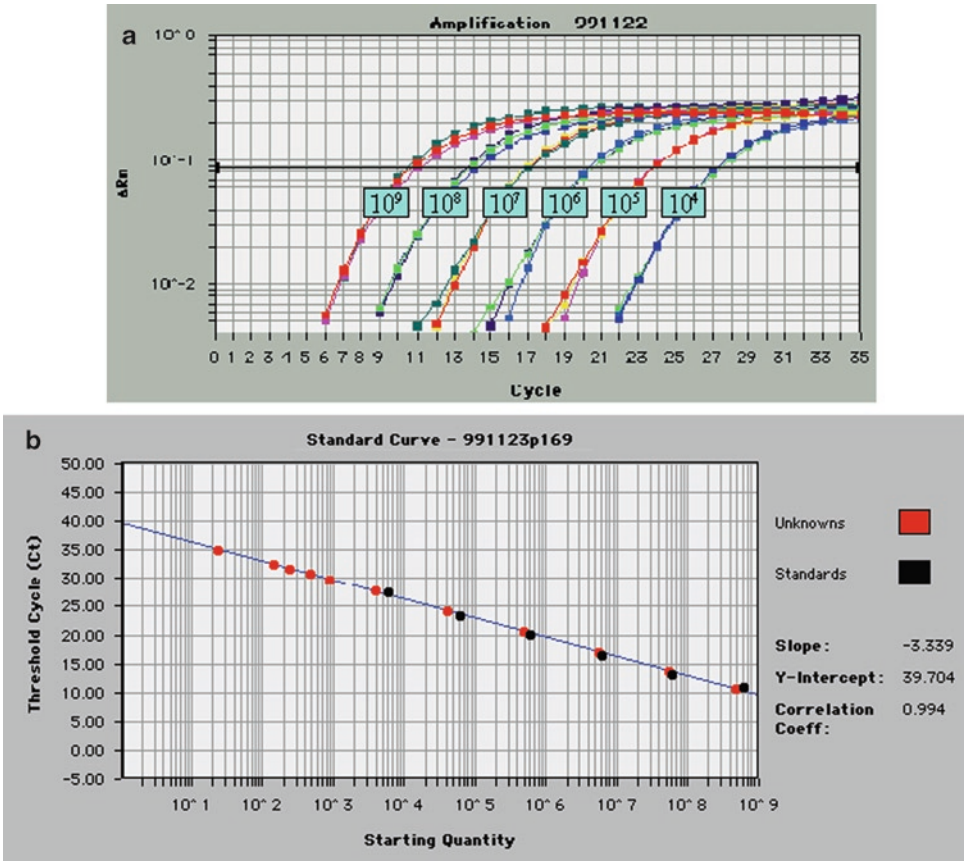


Fig. 3. Real-time asynchronous PCR using PNA probes: (**a**) Amplification curves; (**b**) Standard curve – correlation between the DNA copy number input and Ct (unknown samples represent experimental test samples). DNA input ranges from 10^4 to 10^9 copies per PCR in a tenfold dilution series of LIG1 target. The cycling parameters are described in Fig. 2. and Subheading 3.2.

low and variable yield of PCR products. Robust single-stranded PCR products can be accomplished using modified aPCR (see Note 8). Detailed demonstration and verification of single-stranded PCR product are shown in Fig. 4a–d.

3.3.1. DNA Amplification and Labeling

1. Targets are amplified by aPCR in an ABI 9700 Thermocycler. The forward primer is designed to have a higher Tm of 74°C with Cy5 or Cy3 dye attached to the 5′-end. The reverse primer has a lower Tm and no label.
2. The PCRs are performed in 50 μL containing 1× AmpliTaq Gold® 360 Master Mix, 0.9 μM Cy5- or Cy3-labeled forward primer, 0.1 μM unlabeled reverse primer, and DNA targets in a 96-well PCR plate.
3. The PCR protocol includes two distinctive cycling stages. The first cycling stage consists of an initial 10 min denaturation

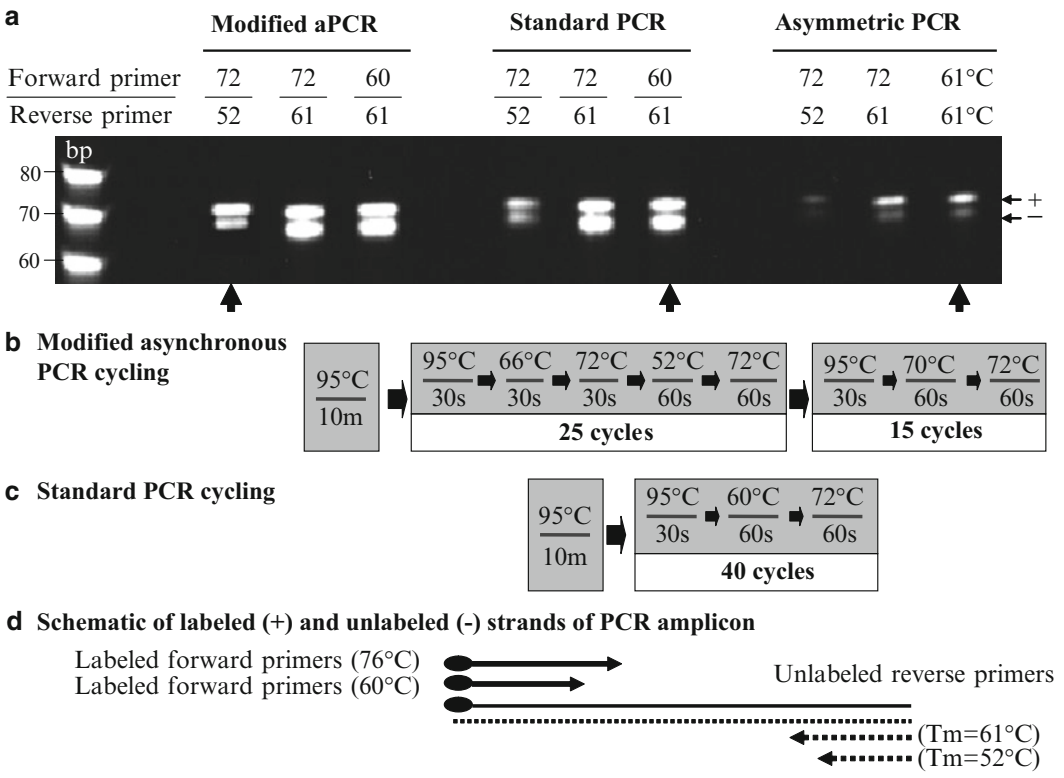


Fig. 4. Comparison of traditional PCR, asymmetric PCR, and modified asynchronous PCR (aPCR) for the generation of LIG1 single-stranded DNA targets: (**a**) Polyacrylamide (15%) gel electrophoresis analysis under denaturing conditions (7 M urea at 60°C) and SYBR® Green II staining of amplicons for three PCR protocols: standard, asymmetric, and modified aPCR. Each PCR protocol uses a combination of forward and reverse primers with different Tm and concentrations. Forward primers are labeled with biotin at 5′. PCR amplicons become single-stranded under denaturing conditions. The upper DNA strand (+) moves slowly because of the biotin label. Results suggest that aPCR generates much higher and reproducible single-stranded PCR products. (**b**) Cycling protocol for modified aPCR includes 25 cycles of aPCR cycling and additional five cycles of highly stringent PCR at 70°C annealing and extension. High temperature annealing greatly inhibits PCR priming from reverse primers, resulting in the production of excessive + strand PCR products. (**c**) Standard PCR cycling for both traditional and asymmetric PCR. (**d**) Schematic illustration of biotin-labeled primers for aPCR, asymmetric PCR, and standard PCR.

at 95°C followed by 25 cycles [95°C for 30 s, 66°C for 30 s (for forward priming), 72°C for 30 s, 52.5°C for 60 s (for reverse priming), and 72°C for 60 s]. The second cycling stage promotes the production of dye-labeled ssDNA. It consists of 15 cycles [95°C for 30 s, 70°C for 60 s, and 72°C for 60 s].

3.3.2. Purification of PCR Products

PCR products are purified to remove excessive labeled primers in three washes on a Microcon YM-100 according to the manufacturer's protocol.

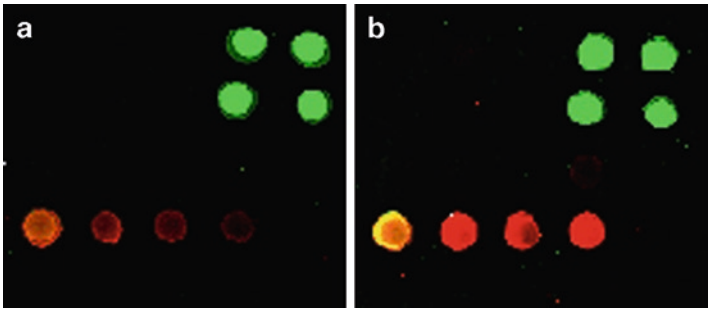


Fig. 5. Array hybridization of Cy5/3-labeled PCR products: (**a**) double-stranded DNA targets produced by traditional PCR protocol; (**b**) predominantly single-stranded DNA targets produced by asynchronous PCR protocol. This figure illustrates the hybridization result of Cy5/3 end-labeled aPCR (mostly ssDNA) and traditional PCR (dsDNA products) from four different targets to a glass slide array. What would appear as light (green) and dark (red) spots represent Cy3- and Cy5-labeled products, respectively. Four spots located at the *upper right corner* are positive hybridization controls. The *bottom* four spots represent hybridization signals to four synthetic targets at different concentrations, high (*left*) to low (*right*) spiking in PCR or aPCRs from high to low. The averaged median signal from the labeled aPCR products was three- to fourfolds higher than that from regular PCR products. The results suggest that the array probes attached to a glass surface hybridize to ssDNA targets more effectively.

3.3.3. Microarray Hybridization, Washing, Data Collection, and Analysis

1. Hybridization mixture (25 μL per slide) consists of microarray hybridization buffer and 1–2 μL PCR products. The mixture is denatured at 95°C for 3 min and 25 μL applied to each slide.
2. The slides are placed inside an array chamber and incubated at 55°C in a water bath for 18 h.
3. After hybridization, wash briefly in the Microarray Wash Buffer I at 55°C.
4. Wash once for 2 min in the Microarray Wash Buffer II at room temperature.
5. Finally, wash twice in the Microarray Wash Buffer III at room temperature for 2 min each.
6. Microarray slides are imaged using an Axon scanner, and images are analyzed in GenePix Pro 3.0 software (Fig. 5).

4. Notes

1. Modifications of PNA and TaqMan probes include:
   MGB = minor grove binder
   6-FAM = 6-carboxyfluorescein
   Flu = fluorescein
   Glu = glutamic acid
   Lys = lysine
   Dabcyl = 4-(4-dimethylaminophenyl) diazenylbenzoic acid
   NH_2 = amine

2. Quantitation of genomic DNA or cDNA using The TaqMan® RNase P Detection Reagents: it is important to use the same and appropriate amount of DNA for all samples. For RNA samples, use the A260 value to determine the appropriate input into the cDNA synthesis reaction and assume 100% conversion into cDNA, then dilute to 10–50 ng/μL. For genomic DNA samples, it is highly recommended to use: (1) UV absorbance (A260/A280) measurements to determine initial DNA concentration and then dilute to 50–100 ng/μL; (2) the TaqMan® RNase P Detection Reagents to create standard curves with DNA Templates for use in qPCR (for details please see the protocol at http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_042486.pdf).
3. Method to visualize both + and – strands of PCR amplicons: Molecular Probes' SYBR® Green II Gel Stain Solution is used to visualize single-stranded PCR amplicons. SYBR® Green II is one of the most sensitive dyes known for detecting single-stranded DNA (ssDNA) in electrophoretic gels. Up to 100 pg ssDNA per band can be detected in a SYBR® Green II stained polyacrylamide gel according to the manufacturer's protocol (http://probes.invitrogen.com/media/pis/mp07568.pdf).
4. Design standard PCR and aPCR primers and probes: The Primer Express® Software Version 2.0 is used for designing qPCR primers and probes. The software requires the target sequence, primer reaction concentration, and desired primer melting temperature (Tm) to select the appropriate sequences. The default settings give the most applicable designs for standard TaqMan assays. For aPCR applications, the forward primer Tm should be set to 74°C and the reverse primers to 54°C. Forward primers with Tm between 72 and 76°C and the reverse primers between 52 and 56°C are recommended.
5. Relative standard curve method: relative DNA copy number can be estimated based on a standard curve of a calibrator. It often requires a series of dilutions in several logs of the calibrator. Use of the gene or target-specific standard curves for better estimates of quantities is highly recommended.
6. Method used for gel electrophoresis and image analysis of PCR products: typically, 5 μL of the PCR product is mixed with a final concentration of 1× loading buffer (45 mM Tris base, 45 mM boric acid, 0.4 mM EDTA, 3% Ficoll, 0.02% bromophenol blue, 0.02% xylene cyanol, pH 8.3) and denatured at 95°C for 15 min. The sample is loaded into a 10–15% denaturing PAGE gel and run in 1× TBE (89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8.3) at 130 V, 70°C for 40 min. The extended product is visualized by staining the gel with 1× SYBR® Green II Staining Solution in a volume of

80 mL in 1× TBE for 20 min. The image is captured in a ChemiImaging 2000 gel documentation system. The amount of DNA can be quantified using SpotDenso program.

7. To visualize both + and – strands of PCR products, forward PCR primers are labeled with a biotin moiety at the 5′-end. PCR products (+ strands) contain a biotin moiety and move slowly (Fig. 4d). Because reverse primers are unlabelled, the PCR products (– strands) move relatively fast. Therefore, differential labeling of the + and – strands separates + from – strands in a denaturing PAGE gel (Fig. 4a).
8. We found that aPCR could generate more robust ssDNA targets than asymmetric PCR. Modified aPCR produces a high yield of single-stranded PCR products, due to difference in Tm and concentrations between forward and reverse primers. Asymmetric PCR generates a low and variable yield of single-stranded PCR products, primarily due to low concentration of the reverse primer (10 nM) and often low PCR efficiency.

References

1. Holland, P. M., Abramson, R. D., Watson, R., and Gelfand, D. H. (1991) Detection of specific polymerase chain reaction product by utilizing the 5'–3' exonuclease activity of Thermus aquaticus DNA polymerase, *Proc Natl Acad Sci U S A* **88**, 7276–80.
2. Higuchi, R., Dollinger, G., Walsh, P. S., and Griffith, R. (1992) Simultaneous amplification and detection of specific DNA sequences, *Biotechnology (N Y)* **10**, 413–7.
3. Wittwer, C. T., Herrmann, M. G., Moss, A. A., and Rasmussen, R. P. (1997) Continuous fluorescence monitoring of rapid cycle DNA amplification, *Biotechniques* **22**, 130–8.
4. Whitcombe, D., Theaker, J., Guy, S. P., Brown, T., and Little, S. (1999) Detection of PCR products using self-probing amplicons and fluorescence, *Nat Biotechnol* **17**, 804–7.
5. Nazarenko, I., Lowe, B., Darfler, M., Ikonomi, P., Schuster, D., and Rashtchian, A. (2002) Multiplex quantitative PCR using self-quenched primers labeled with a single fluorophore, *Nucleic Acids Res* **30**, e37.
6. Nazarenko, I. A., Bhatnagar, S. K., and Hohman, R. J. (1997) A closed tube format for amplification and detection of DNA based on energy transfer, *Nucleic Acids Res* **25**, 2516–21.
7. Gibson, U. E., Heid, C. A., and Williams, P. M. (1996) A novel method for real time quantitative RT-PCR, *Genome Res* **6**, 995–1001.
8. Kutyavin, I. V., Afonina, I. A., Mills, A., Gorn, V. V., Lukhtanov, E. A., Belousov, E. S., Singer, M. J., Walburger, D. K., Lokhov, S. G., Gall, A. A., Dempcy, R., Reed, M. W., Meyer, R. B., and Hedgpeth, J. (2000) 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures, *Nucleic Acids Res* **28**, 655–61.
9. Livak, K. J., Flood, S. J., Marmaro, J., Giusti, W., and Deetz, K. (1995) Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization, *PCR Methods Appl* **4**, 357–62.
10. Fiandaca, M. J., Hyldig-Nielsen, J. J., Gildea, B. D., and Coull, J. M. (2001) Self-reporting PNA/DNA primers for PCR analysis, *Genome Res* **11**, 609–13.
11. Lee, L. G., Connell, C. R., and Bloch, W. (1993) Allelic discrimination by nick-translation PCR with fluorogenic probes, *Nucleic Acids Res* **21**, 3761–6.
12. Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S. M., Driver, D. A., Berg, R. H., Kim, S. K., Norden, B., and Nielsen, P. E. (1993) PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules, *Nature* **365**, 566–8.
13. Svanvik, N., Nygren, J., Westman, G., and Kubista, M. (2001) Free-probe fluorescence of light-up probes, *J Am Chem Soc* **123**, 803–9.

14. Wolffs, P., Knutsson, R., Sjoback, R., and Radstrom, P. (2001) PNA-based light-up probes for real-time detection of sequence-specific PCR products, *Biotechniques* **31**, 766–71.
15. Tyagi, S. and Kramer, F. R. (1996) Molecular beacons: probes that fluoresce upon hybridization, *Nat Biotechnol* **14**, 303–8.
16. French, D. J., Archard, C. L., Brown, T., and McDowell, D. G. (2001) HyBeacon probes: a new tool for DNA sequence detection and allele discrimination, *Mol Cell Probes* **15**, 363–74.
17. Afonina, I. A., Reed, M. W., Lusby, E., Shishkina, I. G., and Belousov, Y. S. (2002) Minor groove binder-conjugated DNA probes for quantitative DNA detection by hybridization-triggered fluorescence, *Biotechniques* **32**, 940–9.
18. Gyllensten, U. B. and Erlich, H. A. (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus, *Proc Natl Acad Sci U S A* **85**, 7652–6.

Chapter 17

Novel Applications of PCR Through the Use of DNA Substrates

Stuart M. Wilson

Abstract

The ability to amplify genetic material using PCR has transformed the field of diagnostics. Now any organism can be detected by identifying the presence of specific nucleic acids. However, there still remain areas in which traditional PCR cannot easily be applied. In this chapter, we describe a different flavour of PCR in which an enzyme acts on a DNA substrate to convert it into a DNA product that can be detected by PCR. This opens up new areas for the application of the PCR technique and we will show how the technique can be used to detect immunoconjugates, bacterial drug resistance, and bacterial contamination.

Key words: Polymerase chain reaction, Bacterial contamination, Drug resistance testing

1. Introduction

1.1. The Limitations of Traditional PCR

Traditionally, amplification methods such as the Polymerase Chain Reaction have been used as a tool to amplify the nucleic acid of interest from an organism. Nucleic acid amplification has an important function in a diverse field of applications including; diagnostics, drug resistance studies, cloning, and typing (1). In diagnostics, organism-specific gene sequences are amplified and detected in order to establish the presence of the organism. Using PCR, for example, it is possible to detect very small numbers of an organism in a sample. It is clear that a highly sensitive detection can be performed if the target is nucleic acid. But not all targets of interest are nucleic acids. For example, in cancer it may be desirable to detect certain cancer-specific marker proteins. There are also a whole spectrum of diseases termed "Protein Aggregation Diseases," that include, for example, Creutzfeldt–Jakob Disease,

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_17, © Springer Science+Business Media, LLC 2011

Huntington's Disease, Alzheimer's Disease, and Parkinson's Disease in which the marker of disease is an abnormal protein. Even though there is a requirement for a sensitive test that can detect the abnormal protein in blood, direct PCR cannot be used for diagnosis as there is no nucleic acid associated with the disease. Immunodetection methods such as ELISA (enzyme-linked immunosorbent assay) are the poor relation to PCR when comparing sensitivity. Even with the best labels on the detection antibodies the sensitivities of these tests are at the picogram levels which correspond to 10^6–10^8 copies of the target. Numerous studies have attempted to apply the sensitivity of PCR to immunodetection usually by binding a PCR-detectable marker (nucleic acid) to the detection antibody – so-called immuno-PCR. Although this technique can achieve high sensitivity in the research laboratory there are still many problems in translating this technique to a robust commercial product (2).

Another area in which PCR cannot easily be applied is as a generic test for the presence of bacteria. In certain situations, in sepsis or platelet products, for example, there is a desire to know whether any bacteria are present at all. For example, platelets are collected and stored at room temperature for up to 5 days. This gives ample opportunity for contaminating bacteria to multiply to dangerous levels which may not have been picked up by culture which is performed at day 1 (3). PCR is good at targeting specific gene sequences but it is difficult to identify sequences that are present in all bacteria and that can be used as PCR targets. Multiplex PCR tests that identify many organisms have been developed (4) but it is difficult to ensure that such a test performs equally well on all of the bacterial targets. In addition, there is still an issue with sensitivity even with PCR. PCR can only detect the organism if the nucleic acid target is actually present in the reaction. This is an issue of sampling – what volume and proportion of the prepared sample is actually analysed by PCR?

1.2. An Alternative Approach

To deal with the limitations of PCR, namely; inability to detect non-nucleic acid targets and lack of sensitivity we have developed an alternative approach. This approach involves the use of nucleic acid as a substrate. These nucleic acid substrates undergo conversion to a product that can be detected by a subsequent PCR. In all examples below, the conversion to PCR-detectable product is catalysed by the presence of an enzyme. This enzyme can be the enzyme component of an enzyme–antibody conjugate; an enzyme present in an organism or an enzyme activated by a co-factor that is present in an organism. Examples and applications of all of these approaches are given in this chapter.

2. Materials

For all methods a PCR machine and hot-start PCR mix is required. We have used a Chromo4 real-time PCR machine (MJ Research) in conjunction with PCR mix (RT-SN2X-03+NR, Eurogentec) containing SYBR Green I.

2.1. Use of Nucleic Acid Substrate for the Detection of Alkaline Phosphatase

1. Calf intestinal alkaline phosphatase (M02090, New England Biolabs, Ltd.).
2. 50 mM Tris–HCl, 10 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT pH 7.5.
3. AP Oligo 1, 5′ GCC GAT ATC GGA CAA CGG CCG AAC TGG GAA. GGC GCA CGG AGA GAC CAC G with 3′ phosphate. AP Oligo 2, 5′ TAG GCG TCG GTG ACA AAC GGC CAG CTA TGA CTT CGT GGT CTC TCC GTG with 3′ phosphate.
4. PCR primers: APF (forward), 5′ GGA CAA CGG CCG AAC TGG GAA G GCG 3′ and APR (reverse), 5′ TAG GCG TCG GTG ACA AAC GGC CAG C 3′
5. Agarose and agarose gel running buffer.

2.2. Use of Nucleic Acid Substrate for the Detection of ATP

1. T4 DNA ligase (M0202, 400,000 units/ml, New England Biolabs, Ltd.).
2. 50 mM HEPES-NaOH, pH 7.5, 5 mM sodium pyrophosphate.
3. 50 mM HEPES-NaOH, pH 7.5, 10 mM dithiothreitol, 1 mM $MgCl_2$.
4. 10 mM Tris–HCl pH 7.4, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 50% glycerol.
5. Tris-buffered saline (TBS) which is 50 mM Tris–HCl, 150 mM NaCl pH 7.5.
6. TBS, 1 mM $MgCl_2$.
7. ATP Oligo 1: 5′ GCC GAT ATC GGA CAA CGG CCG AAC TGG GAA GGC GCA CGG AGA GA 3′, ATP Oligo 2: 5′CCA CGA AGT ACT AGC TGG CCG TTT GTC ACC GAC GCC TA 3′ both with a 5′ and a 3′ phosphate and ATP Oligo 3: 5′ TAG TAC TTC GTG GTC TCT CCG TGC 3′ with a 3′ phosphate.
8. PCR primers ATPF; 5′ GGA CAA CGG CCG AAC TGG GAA GGC G 3′ and ATPR, 5′ TAG GCG TCG GTG ACA AAC GGC CAG C 3′.

2.3. Use of the ATP Assay for Antibiotic Susceptibility Testing of Cultured Bacteria

1. *Staphylococcus aureus* (ATCC strain number 25923).
2. Methicillin-resistant *Staphylococcus aureus* (MRSA).
3. Mueller Hinton (70192, Sigma-Aldrich Ltd.) nutrient broth culture medium.
4. 80 mM NaOH, 1% (v/v) Triton X-100.
5. 80 mM HCl.
6. 50 mM Tris–HCl pH 8.0, 1 mM $MgCl_2$.
7. Oligonucleotides and PCR primers as described in Subheading 2.2.

2.4. Use of Nucleic Acid Substrate for the Detection of Bacterial Ligase

1. NAD Oligo 1, 5′GCC GAT ATC GGA CAA CGG CCG AAC TGG GAA GGC GCA CGG AGA GA 3′; NAD Oligo 2, 5′CCA CGA AGT ACT AGC TGG CCG TTT GTC ACC GAC GCC TA 3′; NAD Oligo 3, 5′ TAG GCG TCG GTG ACA AACGGCCAGCTA GTA CTT 3′; NAD Oligo 4, 5′ CGT GGT CTC TCC GTG CGC CTT CCC AGT TCG GCC GTT GTC CGA TAT3′. NAD Oligo 2 and NAD Oligo 4 both have a 5′ and a 3′ phosphate.
2. Platelets were purchased from the Nation Blood Centre, London, UK.
3. 500 mM Sodium carbonate–sodium hydrogen carbonate, 1% (w/v) *N*-lauroyl sarcosine pH 9.6.
4. Phosphate-buffered saline (PBS), which is 10 mM sodium phosphate, 150 mM NaCl, pH 7.5.
5. PBS, 1 mM EDTA, 4 mg/ml lysostaphin (L0761, Sigma-Aldrich Ltd.).
6. 50 mM Tris–HCl pH 8.0, 1 mM $MgCl_2$, 1 mM dithiothreitol.
7. PCR primers, NADF, 5′ GGA CAA CGG CCG AAC TGG GAA GGC G 3′ and NADR, 5′ TAG GCG TCG GTG ACA AAC GGC CAG C 3′.

3. Methods

3.1. Use of Nucleic Acid Substrate for the Detection of Alkaline Phosphatase

The two mostly commonly used enzymes in antibody conjugates are horseradish peroxidase and alkaline phosphatase. The latter enzyme can be used as an enzyme that converts a nucleic acid substrate into a nucleic acid product that can be readily detected by PCR. The strategy is demonstrated in Fig. 1. It involves the use of two overlapping synthetic oligonucleotide strands that are blocked at the 3′ ends by phosphate groups. With the 3′ phosphate groups in place, these oligonucleotides cannot be extended by nucleic acid polymerase. However, if the blocking phosphate groups at one or more 3′ ends are removed by phosphatase, then

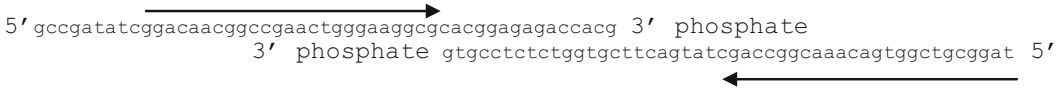


Fig. 1. This shows the design of the DNA substrate for detection of alkaline phosphatase. The two oligonucleotides overlap by 15 bases and are blocked for extension at the 3′ end by phosphate groups. Removal of any of these 3′ phosphate groups by phosphatase enzyme allows subsequent extension of the corresponding oligonucleotide which generates a template for PCR using primers whose positions are indicated by *arrows above* and *below* the sequence.

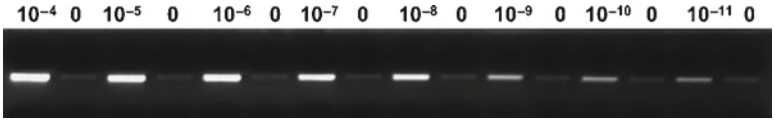


Fig. 2. Titration of alkaline phosphatase and analysis of PCR products. The number of units of alkaline phosphatase used are shown above the relevant lanes.

the relevant oligonucleotide can be extended by nucleotide polymerase which generates a product. With appropriate PCR primer binding sites, this product can be amplified by PCR. Such a nucleic acid substrate can be used to detect phosphatase that is immobilized or free in solution.

1. Tenfold dilutions of calf intestinal alkaline phosphatase were prepared in 50 µl of 50 mM Tris–HCl, 10 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT pH 7.5 containing 1 pmol each of the synthetic oligonucleotides; AP Oligo 1 and AP Oligo 2 and incubated for 60 min at room temperature.
2. After incubation, 10 µl of the solution were analysed by standard hot-start PCR in a volume of 50 µl. PCR conditions were: 1 cycle of 94°C for 15 min; 5 cycles of 94°C for 30 s, 50°C for 30 s, 72°C for 30 s (to allow extension of dephosphorylated oligonucleotides); followed by 25 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 30 s. The PCR primers, APF and APR were used in the PCR (see Note 1).
3. After PCR, 10 µl of reaction were analysed by standard agarose gel electrophoresis (see the results, Fig. 2). In this figure, a negative control containing no phosphatase was included between each phosphatase dilution). The units of alkaline phosphatase used in each reaction shown above the relevant lanes. A titration of signal with increasing dilutions of phosphatise can be seen. In this example, the detection limit for phosphatise was 10^{-11} units which equates to about 60 molecules of alkaline phosphatase (see Note 2, 3 and 4).

3.2. Use of Nucleic Acid Substrate for the Detection of ATP

Detection of bacterial growth is used in the diagnosis and treatment of infectious disease, blood screening, food safety, product quality assurance, and life science research. The measurement of intracellular ATP content has long been the standard for rapid bacterial growth and viability measurement (5–7). The current

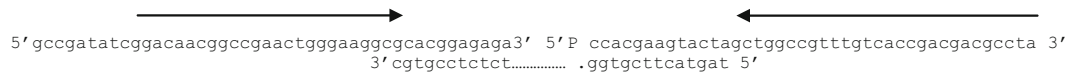


Fig. 3. This shows the design of the DNA substrate for detection of ATP. Two oligonucleotides are juxtaposed through a complimentary oligonucleotide and they can be joined by ligation to form a single long oligonucleotide which generates a template for PCR using primers whose positions are indicated by *arrows above* the sequence. P in the figure represents a 5′ phosphate.

methods used to detect ATP include luminescence generated by the enzyme system firefly luciferase/luciferin. The chemistry involved in this process is simple and can be used with a wide range of luminescence equipment, from handheld devices to sophisticated, laboratory-based instruments for high-throughput applications (8). However, for some applications the sensitivity of this approach may be too low. We have developed an alternative, nucleic acid substrate approach that can be used for detection of ATP in the presence of a DNA ligase, an ATP-requiring enzyme. The enzyme uses ATP to ligate two oligonucleotides in the DNA substrate to create a product that can be amplified in a DNA amplification reaction (Fig. 3).

3.2.1. Preparation of Deadenylated Ligase

Commercial ligase is already charged with one molecule of co-factor which would allow one ligation event if it were not removed by deadenylation with pyrophosphate.

For convenience, this deadenylation can be performed by dialysis.

1. 40,000 U of T4 DNA ligase (400,000 units/ml) were deadenylated by dialysis at 8°C for 1 h against 1 l of 50 mM HEPES, pH 7.5, 5 mM sodium pyrophosphate and then overnight against the same volume of the same buffer.
2. Finally, the ligase was dialysed against a litre of 50 mM HEPES, pH 7.5, 10 mM dithiothreitol, 1 mM $MgCl_2$ for 1 h. This step was repeated.
3. The deadenylated ligase was stored at –20°C in 10 mM Tris–HCl pH 7.4, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 50% glycerol.

3.2.2. Preparation of DNA Substrate

Nicked double-stranded DNA substrate was prepared by heating a mixture containing 10 ng/µl of each of the three oligonucleotides ATP Oligo 1, ATP Oligo 2, and ATP Oligo 3 to 95°C for 5 min and cooling to room temperature.

3.2.3. Detection of ATP

1. 20 µl Reaction volumes containing serial dilutions of ATP were prepared in TBS, 1 mM $MgCl_2$ containing 5 µl of deadenylated ligase, prepared as described above, and 1 µl (10 ng of each oligonucleotide) of the preformed nicked-DNA substrate.

2. After 15 min at room temperature to allow ligation, 5 μl of each reaction was analyzed by PCR in a 50-μl volume with PCR primers ATPF and ATPR. The double-stranded DNA intercalator dye SYBR Green I was present in the PCR master mix to enable detection of amplicon generation in real time. The PCR parameters were 40 cycles of 94°C for 10 s, 65°C for 10 s, and 72°C for 10 s.
3. An example of using the method to detect ATP is shown in Table 1. The results are shown as the Ct0–Ctx where Ct0 is the cycle at which the no-ATP control became positive and Ctx is the cycle at which the test sample became positive. The detection limit under these conditions is 0.5 nmol/L (10 fmol) ATP per 20 μl reaction. The signal generated in the PCR process was proportional to the amount of ATP present in the sample with a dynamic range of detection of at least 100,000-fold (i.e. under the dilutions tested, from 10 fmol to 1 nmol).

3.3. Use of the ATP Assay for Antibiotic Susceptibility Testing of Cultured Bacteria

Bacterial cells contain ATP which declines rapidly with loss of viability. This bacterial ATP can be released from the bacteria and measured using ligase and DNA substrate. For antibiotic susceptibility testing, the assay is used to monitor the bacterial ATP load of a culture grown in the presence and absence of antibiotic. In the absence of antibiotic, the number of organisms and ATP in those organisms should increase with time. However, in the culture with antibiotic, if the organism is sensitive to the antibiotic, there should be little or no increase in numbers of organism or ATP. Conversely, if the bacteria is resistant to drug, there should be growth and an increase in measurable ATP both with and without

Table 1
Limits of detection of ATP

Quantity of ATP in the assay	Ct0–Ctx
1 nmol	14
100 pmol	12
10 pmol	10
1 pmol	7
100 fmol	5
10 fmol	3
0	NA

Ct0 is the cycle at which the no-ATP control became positive and Ctx is the cycle at which the test sample became positive

antibiotic. To convert PCR signal to a figure that is indicative of growth, the Growth Index is used. To calculate the growth index, it is necessary to calculate $2^{(Ct0-Ct3)}$. CT0 is the cycle at which the PCR becomes positive from the culture before growth, i.e. the no-growth control and Ct3 is the cycle at which the PCR becomes positive from a given culture after 3 h growth. This equation is derived from the fact that there is a doubling in the number of PCR amplicons at each cycle so, for example, a cycle difference of four actually represents, at maximum PCR efficiency, $2 \times 2 \times 2 \times 2 = 16$-fold difference (all other parameters being equal) in the number of amplicons and by inference, of target molecules that were present at the two time points prior to cycling which in turn is directly related to the relative numbers of bacilli present at the two time points.

1. *S. aureus* (ATCC strain number 25923) and a MRSA strain of *S. aureus* were grown in Mueller Hinton nutrient broth culture medium until stationary phase.
2. A 1,000-fold dilution of these cultures were made in nutrient broth with and without oxacillin antibiotic at various concentrations. One dilution with no antibiotic was kept at 4°C as a "no-growth" control whereas the other dilutions were incubated at 37°C.
3. After 3 h all cultures including the "no-growth" controls were tested in the ATP assay.
4. Ten microlitre of each culture was treated with 10 µl of lysis solution containing 80 mM NaOH, 1% (v/v) Triton X-100 and heating to 95°C for 3 min.
5. After cooling to room temperature, the mixture was neutralized with 10 µl of 80 mM HCl and incubated with 100 µl ligase mix containing 50 mM Tris–HCl pH 8.0, 1 mM $MgCl_2$, and 5 µl of deadenylated ligase and 10 µl of the preformed nicked-DNA substrate, both prepared as described earlier.
6. After 30 min 5 µl of the mix was analysed by real-time PCR as described earlier.
7. An example of the method is shown in Fig. 4. Although the two bacilli had different growth rates it can be seen that both bacilli strains grew in the absence of drug. When antibiotic at either concentration was present there was no growth of the sensitive bacterial strain but the strain resistant to oxacillin grew at a rate comparable to the no-drug control.

3.4. Use of Nucleic Acid Substrate for the Detection of Bacterial Ligase

Bacterial contamination of platelets is the most common cause of adverse reaction and death due to transfusion-associated infections (3). Screening for contamination is performed by culture. However, culture only quality assures the platelets at the day of culture inoculation and misses some contamination. Particularly

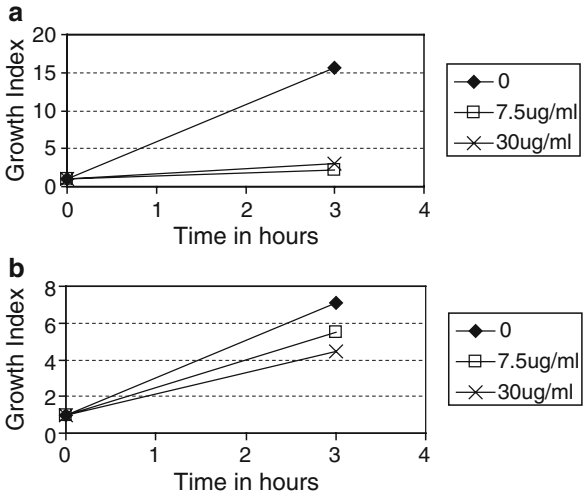


Fig. 4. Strains of a drug-sensitive (**a**) and -resistant (**b**) *S. aureus* were grown for 3 h in the presence and absence of oxacillin and then tested for ATP content. The PCR signal was converted to Growth Index to reflect the growth of the organism which was plotted against time.

problematic and of highest risk of contamination are stored aged platelets – a consequence of the need to store platelets at room temperature. Generic PCR tests have been developed based on the detection of conserved regions of bacterial genes but it is hard to develop such tests that can pick up all bacteria with equal sensitivity. An alternative approach is to detect the NAD-dependent DNA ligase enzyme that is present in all bacteria tested to date. The enzyme can be detected in a simple manner by lysing the bacilli and adding DNA substrate which, in the presence of the bacterial ligase, is ligated into a form that can be detected by a subsequent PCR. The high sensitivity of the method is due to a double cascade of amplification. Firstly, each molecule of bacterial ligase performs multiple ligations and secondly, in the PCR there is amplification of these ligated products.

In this example, platelets are spiked with dilutions of bacilli and the presence of the contaminating bacilli detected by the bacterial NAD-dependent DNA ligase activity.

3.4.1. Preparation of DNA Substrate

1. Double-stranded DNA substrate with a 5-bp overhang (see Fig. 5) was prepared by heating a mixture containing 10 ng/µl of each of the four oligonucleotides NAD Oligo1, NAD Oligo 2, NAD Oligo 3, and NAD Oligo 4, to 95°C for 5 min and cooling to room temperature.

3.4.2. Testing of Platelets for Bacterial Contamination

1. *S. aureus* (ATCC strain number 25923) was grown in Mueller Hinton nutrient broth culture medium until stationary phase. The numbers of bacilli were counted by plating out dilutions on nutrient agar and counting colonies.

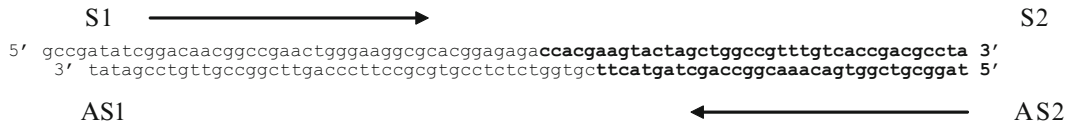


Fig. 5. The DNA substrate for detection of NAD-dependent DNA ligase consists of four oligonucleotides; S1, S2, AS, and AS2 with a 5-bp complementary overlap. S2 and AS have phosphate groups at the 5′ ends to enable ligation. PCR primer positions are indicated by *arrows above* and *below* the sequence.

Table 2
Detection of bacilli spiked into platelets

Number of bacilli in the platelets	Ct0–Ctx
10^6	16
10^5	12
10^4	9
10^3	5
10^2	1
0	NA

The results are shown as the Ct0–Ctx where Ct0 is the cycle at which the no-bacteria control became positive and Ctx is the cycle at which the test sample became positive

2. Serial dilutions of this culture were performed in 0.5 ml platelets in order to mimic contaminated platelets.
3. In order to lyse the platelets 0.5 ml of 500 mM sodium carbonate, 1% (w/v) *N*-lauroyl sarcosine pH 9.6 was added and mixed.
4. Bacilli were pelleted from the lysed platelets by centrifugation at 8,000 × *g* for 5 min and resuspended in 100 µl PBS.
5. After repeat centrifugation the pellets were resuspended in 20 µl of lysis buffer containing PBS, 1 mM EDTA, 4 mg/ml lysostaphin (see Note 4.3.1).
6. After 10 min at 4°C, 2 µl of lysed sample were added to a 20-µl ligation containing 10 ng each oligonucleotide and 50 mM Tris–HCl pH 8.0, 1 mM $MgCl_2$, 1 mM dithiothreitol and incubated for 30 min at room temperature.
7. After ligation, 2 µl of the ligation reaction was analysed by real-time PCR using PCR primers NADF and NADR. The PCR parameters were: 40 cycles of 94°C for 10 s, 65°C for 10 s, and 72°C for 10 s (See Note 5).
8. An example of this method is shown in Table 2. The results are shown as the Ct0–Ctx where Ct0 is the cycle at which the

no-bacteria control became positive and Ctx is the cycle at which the test sample became positive. The detection limit under these conditions is less than 1,000 bacilli spiked into 0.5 ml platelets (See Note 6).

4. Notes

1. Some commercial Taq enzymes may contain phosphatase enzyme as a contaminant. This will cause a background in the assay. To check for this, compare two scenarios; one where the PCR is performed immediately and one where the PCR mix containing reaction product is left for 30 min prior to PCR. If there is a problem with phosphatase contamination, the latter scenario will cause an increased background.
2. If some oligonucleotides in the DNA substrate are not blocked by phosphate at the 3′ end this will lead to a high background. The quality of these oligonucleotides is therefore highly important.
3. The assay has been illustrated with agarose gel analysis of product. In practice real-time PCR is applicable and the incubation steps can be reduced to minimise cycling time. It is advisable to optimise these conditions for each PCR machine.
4. The DNA substrate detects very low quantities of free alkaline phosphatase. The assay performs equally well when used to detect alkaline phosphatase–antibody conjugates. This allows an ELISA type assay to be performed, for example, using immobilised antibodies to capture the antigen followed by detection using alkaline phosphatase–antibody conjugates. After washing, the conjugate itself is detected by adding DNA substrate to the well and detecting its conversion to product by subsequent PCR analysis. In this application, the sensitivity of the method will be determined by the efficiency of the washing step in removing the enzyme conjugate that is not bound to antigen and would thus contribute to background.
5. Some commercial Taq enzymes may contain ligase enzyme as a contaminant. This will cause a background in the assay. To check for this, compare two scenarios: one where the PCR is performed immediately and one where the PCR mix containing reaction product is left for 30 min prior to PCR. If there is a problem with ligase contamination, the latter scenario will cause an increased background after PCR.

6. The assay has applications for highly sensitive and rapid bacterial detection. One of the issues with the method is that lysostaphin is used to break open the Staphyloccocus. This method is not applicable to a wide range of bacteria. Any method that is applied for the detection of platelet contamination must be gentle enough to preserve the bacterial ligase enzyme yet efficient on a wide range of bacterial strains. Recently we have begun investigating bacterial lysis using bead-beating devices such as the FastPrep (Fisher Scientific Ltd., UK).

References

1. Innis, M.A., Gelfand, D.H., Sninsky, J.J, and White, T.J. (Eds.) (1990) PCR Protocols. A guide to methods and application. Academic Press Inc, San Diego, CA.
2. Adler, M., Wacker, R., and Niemeyer, C.M. (2008) Sensitivity by combination: immuno-PCR and related technologies. *Analyst* **133**, 702–18.
3. Eder, A.F., Kennedy, J.M., Dy, B.A., Notari, E.P., Weiss, J.W., Fang, C.T., Wagner, S., Dodd, R.Y., and Benjamin, R.J. (2007) Bacterial screening of apheresis platelets and the residual risk of septic transfusion reactions: the American Red Cross experience (2004–2006). *Transfusion* **47**, 1134–42.
4. Dierkes, C., Ehrenstein, B., Siebig, S., Linde, H.J., Reischl, U., and Salzberger, B. (2009) Clinical impact of a commercially available multiplex PCR system for rapid detection of pathogens in patients with presumed sepsis. *BMC Infect. Dis.* **11**, 126.
5. De Luca, M. and McElroy, W.D. (1978) Purification and properties of firefly luciferase. *Meth. Enzymol.* **57**, 3–15.
6. Kricka, L.J. (1988) Clinical and biochemical applications of luciferases and luciferins. *Anal. Biochem.* **175**, 14–21.
7. Lundin, A., Hasenson, M., Persson, J., and Pousette, A. (1986) Estimation of biomass in growing cell lines by adenosine triphosphate assay. *Meth. Enzymol.* **133**, 27–42.
8. Stanley, P.E. (1992) A survey of more than 90 commercially available luminometers and imaging devices for low-light measurements of chemiluminescence and bioluminescence, including instruments for manual, automatic and specialized operation, for HPLC, LC, GLC and microtitre plates. Part 1: descriptions. *J. Biolumin. Chemilumin.* **7**, 77–108.

Chapter 18

Enhanced Solid Phase PCR for Increased Loading of Amplicon onto Solid Support

Daniel J. Park

Abstract

The loading of amplicons onto solid supports such as beads during multiplex PCR or emulsion PCR conventionally has been performed by use of Solid Phase PCR or asymmetric Solid Phase PCR. These approaches are restrictive with respect to amplification efficiency and degree of amplicon loading. This chapter details Enhanced Solid Phase PCR principles and methodologies to enable higher amplicon loading in the context of uncompromised amplification efficiency.

Key words: Enhanced solid phase PCR, Solid phase PCR, Emulsion PCR, Multiplex PCR

1. Introduction

Often in diagnostic applications and in many emulsion PCR applications, such as relatively recent high-throughput sequencing systems (1–3), very low copy numbers of template nucleic acids are present prior to amplification. In systems where beads have been used as solid supports to facilitate a "read-out," Solid Phase PCR (SP-PCR) (4) or asymmetric Solid Phase PCR (aSP-PCR) (5) have been applied. In SP-PCR, balanced forward and reverse aqueous primers are employed in aqueous PCR in the presence of solid support with attached solid support primer of sequence matching that of one of the aqueous primers. In this setting, aqueous amplification is efficient but loading of amplicons onto solid support is relatively inefficient because the aqueous primer with sequence matching that of the solid support primer out-competes the solid support primer during annealing and extension steps. aSP-PCR restricts the concentration of the aqueous primer with sequence matching that of the solid support

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_18, © Springer Science+Business Media, LLC 2011

primer in an attempt to offer solid support primer a better chance of annealing with amplicon. However, this comes at a cost. A restricted concentration of one of the aqueous primers results in a sub-optimal amplification efficiency and amplicon yield (reduced sensitivity), which in turn impacts on the efficiency of annealing between solid support primer and amplicon (6–8).

Bridge PCR is a form of SP-PCR which uses forward and reverse PCR primers which are physically attached to solid support in the absence of aqueous primers. At each amplification cycle, an amplicon bridges over to anneal with a solid support primer (9, 10). In this method, any amplification which occurs inherently contributes to loading of amplicon onto solid support surface. However, the kinetics of reaction are less efficient than reactions which utilise aqueous primers and as a result, reported sensitivities are low (11, 12).

In this chapter, Enhanced Solid Phase PCR (ESP-PCR) principles and methodologies are presented. ESP-PCR enables higher amplicon loading in the context of uncompromised amplification

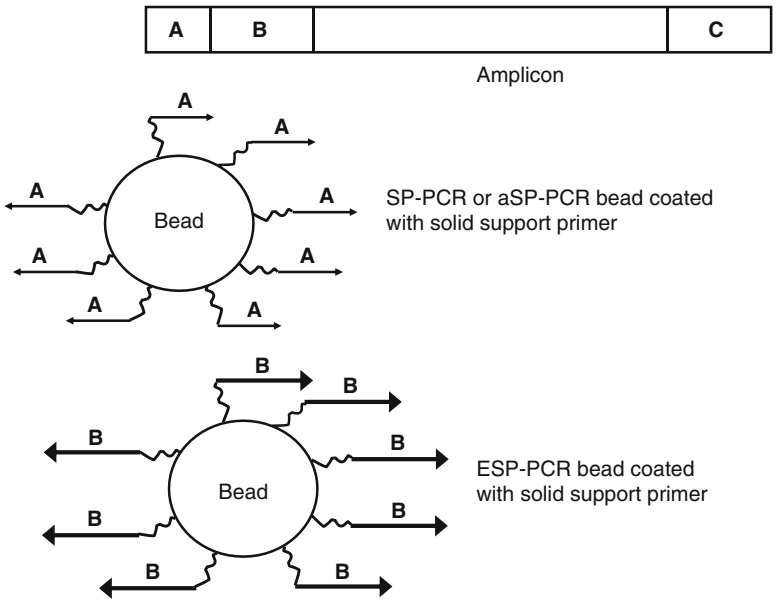


Fig. 1. Schematic illustration of a preferred design of beads used in ESP-PCR. The equivalent reagents of SP-PCR and aSP-PCR are shown alongside for purposes of comparison. The *box* represents an amplicon generated by aqueous primers targeting regions "A" and "C." Region "A" is also targeted by solid support primer in SP-PCR or aSP-PCR. Region "B" ("nested in" from region "A") is targeted by solid support primer in ESP-PCR. In the preferred embodiment of ESP-PCR, bead-conjugated nested solid support primer "B" has a higher Tm than aqueous primer "A." This results in more competitive priming by solid support primer with the option of further facilitating loading of amplicon onto beads by raising the annealing temperature during latter PCR cycles. Aqueous primer "C" also exhibits a relatively high Tm to facilitate amplicon loading onto beads during optional higher annealing temperature steps towards the end of thermocycling.

efficiency (13). ESP-PCR is based on SP-PCR, but matrix-attached solid support primer bears a higher Tm than its aqueous counterpart and/or is "nested in" from its aqueous counterpart. ESP-PCR includes the option of an increased annealing temperature during latter PCR cycles to influence priming competition in favour of the solid support primer (Fig. 1). The adoption of ESP-PCR in emulsion PCR protocols will likely improve the proportion of successful "in-emulsion" amplifications and increase solid support loading with amplicons. In turn, this should require less starting material to template emulsion PCR with implications relating to degree of library amplification required in high-throughput sequencing strategies and concerns over representational bias. Should this transpire, it would represent a possible advantage to the use of emulsion PCR over bridge PCR-based high-throughput sequencing for detection of rarer species from limited starting material (14). Further, higher amplicon loadings should facilitate longer sequencing "read" capability by increasing initial signal-to-noise ratios and thereby allowing more tolerance of petering out of signals during the cycles of sequencing chemistry. In diagnostic applications, beads become useful reporter vehicles without compromising analyte detection sensitivity.

2. Materials

Listed below are ESP-PCR reagents for use in a multiplex PCR, flow cytometry read-out format. The list does not include emulsion PCR materials although considerations relating to emulsion PCR are discussed in Subheading 4 (see Note 1):

1. Template nucleic acid (see Note 2).
2. Hot start DNA-dependent DNA polymerase and accompanying buffers and solutions (see Note 3).
3. 25 or 50 mM magnesium chloride.
4. 10 mM dNTPs.
5. Molecular biology grade water.
6. Forward and reverse aqueous primers designed according to standard PCR primer design considerations, except the reverse primer that is not competitive with solid support primer exhibits a relatively high Tm (see Note 4). In a flow cytometry read-out example, fluorescently labelled reverse aqueous primer (primer "C" with reference to Fig. 1) such as primer bearing 5-prime conjugated Alexafluor647 can be used (see Notes 4 and 5).
7. Beads covalently coated with solid support primer. Preferred solid support primer design in ESP-PCR aims for an amplicon

annealing site "nested in" and with a higher Tm relative to the correspondingly oriented aqueous primer (Fig. 1). Other forms of ESP-PCR include the use of solid support primer that is nested relative to the correspondingly oriented aqueous primer but without significant difference between their Tms and the use of solid support primer which shares some sequence identity with the corresponding aqueous primer but which includes 3-prime extended sequence to result in a higher Tm (see Note 6).

8. Rainbow Calibration Particles (Spherotech) (for flow cytometry applications).

3. Methods

The ESP-PCR protocol will vary depending on the application. The methods detailed in this section relate to those that have been tested in a non-emulsion PCR format with read-out by flow cytometry (13). Alternative ESP-PCR format options are indicated in this section and the accompanying Subheading 4.

3.1. Reaction Set-Up

1. Initiate ESP-PCR set-up per a conventional PCR. For example set up 19 µl standard PCR reagents but with a view to achieving a 20 µl total reaction volume (following microsphere addition in Subheading 3.1, step 2) containing template DNA (in (13) an example using five copies of *Neisseria gonorrhoea* and *Chlamydia trachomatis* DNA against a background of 5 ng Jurkat cell line human genomic DNA is described), 1× PlatinumTaq™ buffer, 2 U PlatinumTaq™ (Invitrogen), 2 mM magnesium chloride, 200 µM dNTPs, and 250 nM aqueous forward and reverse primers (reverse primer, or primer "C" with reference to Fig. 1, labelled at the 5-prime terminus with Alexafluor647) (13) (see Notes 5 and 7).
2. Add 1 µl containing approximately 350 silica microspheres (6.8 µm diameter) of each multiplex type, coated with covalently attached ESP-PCR solid support primers (targeted to position "B" with reference to Fig. 1). Microspheres should be evenly suspended and sedimentation avoided (see Notes 6 and 7). The type and number of beads will vary depending on the application (see Notes 6 and 7).
3. Proceed to thermocycling.

3.2. Thermocycling

Conventional PCR thermocycling conditions may be appropriate in ESP-PCR when the solid support primer is nested with respect to region "A" of Fig. 1 whether it exhibits a higher Tm relative to its aqueous counterpart or not. PCR thermocycling conditions

which increase the annealing temperature in latter cycles could facilitate higher loading of beads with amplicon when the solid support primer exhibits a higher Tm relative to its aqueous counterpart, whether nested or not.

3.2.1. Option 1: Conventional PCR Thermocycling

For use when solid support primer is nested whether it exhibits a higher Tm relative to its aqueous counterpart or not. For example (13):

94°C for 2 min, followed by 50 cycles of (90°C for 30 s, 55°C for 1 min, 72°C for 1 min), followed by 72°C for 5 min.

It should be noted that these conditions were applied in proof-of-concept experiments but it is considered that there should be lots of room for manoeuvre with respect to altering conditions to suit an application. Here, a relatively low denaturation temperature was applied to PCR cycles with a view to minimising degradation of the beads (see Notes 6 and 7).

3.2.2. Option 2: Increased Annealing Temperature During Latter PCR Cycles

For use when the solid support primer exhibits a higher Tm relative to its aqueous counterpart, whether nested or not. One example would be to perform thermocycling per Subheading 3.2.1 but with the final five steps of the 50 cycles conducted with an annealing temperature 5°C higher than the Tm of the competing aqueous primer but not higher than the solid support primer Tm.

3.3. Analysis

1. Wash beads in flow cytometer acquisition buffer using one or multiple spin-wash steps to minimise background fluorescence (see Note 8).
2. Perform flow cytometry using a machine calibrated using fluorescence standards (see Note 9).

4. Notes

1. ESP-PCR has not been tested in the context of emulsion PCR. However, similar principles apply to microdroplet in-emulsion PCR as to conventional larger volume PCR. The size of the beads employed would need to be appropriate to permit a high density of micro-reactors and not destabilize the emulsion. As such, in emulsion PCR protocols, smaller beads than the 6.8 μm diameter beads used in this example chapter would be appropriate. If ESP-PCR design is very constrained by a requirement to minimise amplicon size or by the size of "extraneous" sequence such as library adapters used in high-throughput sequencing methods, solid support primer could include sequence matching that of the corresponding aqueous primer but with the inclusion of additional 3′ sequence to increase the solid support

primer Tm (nested, high Tm solid support primer otherwise preferred). Thermocycling conditions including increased annealing temperatures during latter cycles would then favour loading of beads with amplicon. Amplicons of several hundred base pairs have been used in emulsion PCR regimens (1), but the use of amplicons that are too long may be problematic.

2. The quality of input template nucleic acids in terms of integrity and presence of PCR inhibitors or competitive template species impacts on PCR, generally. In this regard, ESP-PCR is no different. A specific example in the context of high-throughput sequencing is adapter derivatives competing with desired templates. Effort should be made to ensure input template nucleic acids are of the highest quality possible.
3. Hot start thermostable DNA polymerase systems based on antibody or chemical inactivation are preferred for convenience and facilitation of PCR contamination control. Hot start PCR facilitates "clean" reactions with minimization of off-target products.
4. Design of a reverse aqueous primer (primer "C" with reference to Fig. 1) with a higher Tm (approximately 10°C higher is recommended, at a similar Tm to the solid support primer "B" – although there should be considerable scope for design fluctuation based on this idea) than the forward aqueous primer (primer "A" with reference to Fig. 1) facilitates amplicon loading with double-stranded DNA in regimens where increased annealing temperatures during latter cycles are used. In the example used in this chapter, the reverse aqueous primer was also fluorescently labelled with fluorophore and in this context dsDNA amplicon loading onto beads was desirable. Should ssDNA representative loading of amplicon onto beads be preferred, the reverse aqueous primer could be designed with a Tm similar to the forward aqueous primer such that during increased annealing temperatures during latter cycles the bias is strongly towards solely solid support primer primed extension. The reaction could then be stopped chemically or physically prior to a return to conditions permissive to conversion of the ssDNA to dsDNA.
5. The example used in this chapter applied Alexafluor647 covalently conjugated to the 5-prime end of reverse aqueous PCR primer to enable read-out by flow cytometry. The nature of the chemical conjugation must be sufficiently stable in the context of the buffer conditions and temperatures used during PCR. Similarly, the fluorophore must be stable in these conditions. The quality of fluor-labelled oligonucleotide is also very important. Unless a low ratio of fluor-labelled oligo:total

oligo is achieved for this primer, the sensitivity of the system will be compromised.

6. Beads conjugated with solid support primer should be sufficiently stable to withstand prolonged storage and the temperatures required during thermocycling inherent to PCR. Beads should be evenly suspended immediately prior to any transfer steps to negate sedimentation effects. Larger beads tend to sediment at a faster rate than smaller beads. Genera Biosystems covalently conjugate oligos to silica microspheres and use DNA "tether sequence" at the 5-prime end prior to the "oligo proper," to facilitate appropriate presentation of solid support oligos to the aqueous phase. Beads are stored and handled in a non-ionic detergent-containing buffer which is not inhibitory to PCR. The degree of conjugation of oligos onto bead surface can be assessed by hybridizing beads to calibrated, fluorescently labelled oligos bearing sequence that is complimentary to sequence of the bead oligos.
7. The methods presented in this chapter represent proof-of-principle methods. It is envisaged that a wide range of variations on the theme would be tolerated with a view to optimizing conditions for a given application (e.g. nature of thermostable DNA polymerase, bead numbers and size, primer design and cycling parameters). Depending on bead size and sedimentation considerations, master mixes including beads could be prepared during ESP-PCR set-up. Control reactions should be included to account for background signal during read-out. Sedimentation of beads during ESP-PCR thermocycling does not appear to exert a strong influence on the relative loadings of beads with amplicons. Relatively low variation across beads was observed by flow cytometry (13).
8. Washing can be performed using physical phase separation. One option is to apply centrifugation spin-wash steps. If this is performed with silica beads, only relatively gentle centrifugation is required which facilitates rapid resuspension when required. Care should be taken not to aspirate off supernatant with a tip too close to the bead sediment to prevent bead loss. An attractive alternative might be the use of magnetic beads in ESP-PCR such that application of a magnetic field could be used during phase-separation bead washing. However, for flow cytometry read-out, beads must not exhibit excessive background fluorescence, or the sensitivity of the method will be compromised.
9. Flow cytometers should be adjusted to appropriate instrument voltage settings using calibrated fluorescence standards, e.g. Rainbow Calibration Particles. Calibrated fluorescence standard files should also be captured as data files during the

same instrument run using the same settings as controls and test samples. 96-well format compatible flow cytometers such as the FACSArray and FACSCanto (Becton Dickinson) enable convenient specimen handling. Ensure the instrument is instructed to perform a "pipette up and down" resuspension immediately prior to acquisition.

Acknowledgements

This work was funded by Genera Biosystems and undertaken at Genera Biosystems' laboratories.

References

1. Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A. et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.
2. Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., and Church, G.M. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–32.
3. Smith, D.R., Quinlan, A.R., Peckham, H.E., Makowsky, K., Tao, W., Woolf, B. et al. (2008) Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res.* **18**, 1638–42.
4. Adessi, C., Matton, G., Ayala, G., Turcatti, G., Mermod, J.-J., Mayer, P., and Kawashima, E. (2000) Solid phase amplification: characterization of primer attachment and amplification mechanisms. *Nucleic Acids Res.* **28**, e87.
5. Shapero, M.H., Leuther, K.K., Nguyen, A., Scott, M., and Jones, K.W. (2001) SNP genotyping by multiplexed solid-phase amplification and fluorescent minisequencing. *Genome Res.* **11**, 1926–34.
6. Poddar, S.K. (2000) Symmetric vs. asymmetric PCR and molecular beacon probe in the detection of a target gene of adenovirus. *Mol. Cell. Probes* **14**, 25–32.
7. McCabe, P.C. (1999), in: Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. (Eds.), PCR Protocol: A Guide to Methods and Applications, Academic Press, New York, pp. 76–83.
8. Gyllenstein, U.B. and Allen, M. (1993) Sequencing of in vitro amplified DNA. *Methods Enzymol.* **218**, 3–16.
9. Rasmussen, S.R., Rasmussen, H.B., Larsen, M.R., Hoff-Jorgensen, R., and Cano, R.J. (1994) Combined polymerase chain reaction-hybridization microplate assay used to detect bovine leukemia virus and salmonella. *Clin. Chem.* **40**, 200–5.
10. Oroskar, A.A., Rasmussen, S.-E., Rasmussen, H.N., Rasmussen, S.R., Sullivan, B.M., and Johansson, A. (1996) Detection of immobilized amplicons by ELISA-like techniques. *Clin. Chem.* **42**, 1547–55.
11. Tillib, S.V., Strizhkov, B.N., and Mirzabekov, A.D. (2001) Integration of multiple PCR amplifications and DNA mutation analysis by using oligonucleotide microchip. *Anal. Biochem.* **292**, 155–60.
12. Mercier, J.-F., Slater, G.W., and Mayer, P. (2003) Solid phase DNA amplification: a simple Monte Carlo Lattice model. *Biophys. J.* **85**, 2075–86.
13. Khan, Z., Poetter, K., and Park, D.J. (2008) Enhanced solid phase PCR: mechanisms to increase priming by solid support primers. *Anal. Biochem.* **375**, 391–3.
14. Schuster, S.C. (2008) Next-generation sequencing transforms today's biology. *Nat. Methods* **5**, 16–8.

Chapter 19

Application of Blocking Oligonucleotides to Improve Signal-to-Noise Ratio in a PCR

Hege Vestheim, Bruce E. Deagle, and Simon N. Jarman

Abstract

"Universal" or group-specific PCR primers have a tendency to predominately hybridise with the common sequences in samples with mixed templates. The result is that the rarer sequences are seldom retrieved by cloning or sequencing. The use of a blocking oligonucleotide (oligo) designed to specifically prevent amplification of dominant or unwanted DNA templates is an easy way to improve the amplification of rarer sequences. Here, we describe the different types of blocking principles and the different types of blocking oligos and give guidelines and examples of their application.

Key words: Blocking probes, Blocking primers, Blocking oligonucleotide, Competitive probes, C3 spacer, peptide nucleic acids, Locked nucleic acids, Allele-specific competitive blocker PCR, Enhanced amplification of rare sequences

1. Introduction

Many DNA templates used in PCR contain non-target sequences that may be co-amplified along with the target sequences. In some DNA templates, non-target sequences may even be far more prevalent than the sequences of interest. Examples include the presence of parasite DNA in blood samples (1), rare mutations in clinical samples (2), food DNA in dietary samples (3), or rare bacterial sequences in environmental samples (4). It is ideal to use conserved primers to PCR amplify DNA markers from such samples when detailed knowledge of the target sequences is lacking or when the targeted DNA comes from a diverse group of sequences. This is problematic, however, because even after PCR the target DNA will still be rare, or worse, more common sequences may be amplified in early rounds of the PCR and the rarer sequences may not be represented at all in the completed

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_19, © Springer Science+Business Media, LLC 2011

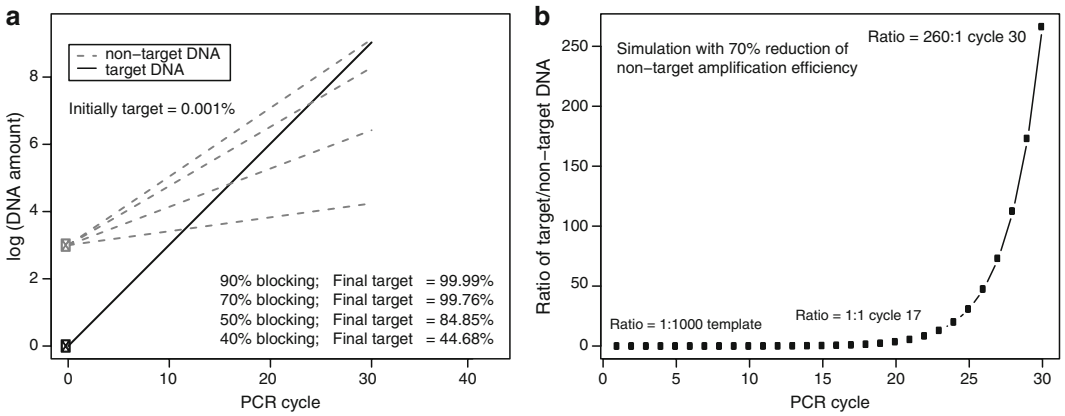


Fig. 1. Effect of a reduction in non-target DNA amplification efficiency on the composition of amplicons during PCR. (**a**) Simulated PCR amplification of target and non-target DNA with an initial ratio of 1:1,000 and reduction of non-target amplification efficiency by 40, 50, 70, or 90% relative to doubling of target DNA during each cycle. Percentage of amplicons from the target after 30 cycles is shown. (**b**) Exponential increase in ratio of target to non-target amplicons during PCR when non-target amplification efficiency is reduced using blocking oligos.

reaction (5). This situation is exemplified by DNA diet studies where DNA template is purified from dietary samples such as stomach contents or faeces (6). These samples tend to be dominated by DNA derived from the digestive tract of the animal being studied rather than by food DNA (7), and the taxonomic diversity of the food species to be identified is usually not known and can be extremely diverse.

The addition of blocking oligos that suppress the amplification of common sequences that are not desired can be a good solution for this problem (8–11). Even when target DNA represents a very low proportion of initial template, a moderately effective blocking oligo (e.g. producing a 70% reduction in amplification efficiency of non-target template) results in an excess of target amplicons after 30 cycles (Fig. 1a). The ratio of target to non-target DNA increases at an exponential rate as PCR proceeds, so increasing the number of PCR cycles will have a large impact on the proportion of target DNA in the completed reaction (Fig. 1b). In this chapter, we discuss the types of blocking oligos that have been used and suggest procedures for their design and application.

2. Materials

2.1. Different Types of Blocking Oligos

2.1.1. Standard DNA Oligos Modified to Not Prime Amplification

There are now many different types of blocking oligos available. One group is *standard DNA oligos modified to not prime amplification*. This is a DNA oligo that is modified so that it does not serve as an initiation point for polymerisation of a complementary strand by *Taq* polymerase or similar enzymes. A number of modifications capable of inactivating the oligo are offered from most

suppliers of custom oligonucleotides including a *3′-Spacer C3 CPG* (1-dimethoxytrityloxy-propanediol-3-succinoyl)-long chain alkylamino-CPG) (11), *3′ phosphate group* (9), and *chemically reversed 3′terminal nucleotide* (3′ to 5′)/*Inverted end* (10).

2.1.2. Peptide Nucleic Acids

Peptide nucleic acids (*PNAs*) may also effectively reduce background PCR amplification (12–16). PNAs are oligonucleotides where the backbone is composed of repeating *N*-(2-aminoethyl)-glycine units linked by peptide bonds. PNA/DNA duplexes have a higher thermal stability (melting temperature) compared with DNA/DNA duplexes because the backbone of PNA contains no charged phosphate groups and there is no electrostatic repulsion (17). A PNA probe will not prime DNA polymerisation because PNA oligomers are not recognised by DNA polymerase (18).

2.1.3. Locked Nucleic Acids

And a third group that can be used as blocking oligos are *locked nucleic acids* (*LNAs*) (13, 19). LNAs are modified RNA nucleotides where the ribose is modified with an extra bridge connecting the 2′ and 4′ carbons. The locked ribose conformation enhances base stacking and backbone pre-organisation. LNAs have also a higher thermal stability than DNA oligonucleotides.

2.2. Special Considerations/ Polymerases

The characteristics of polymerases for use with blocking oligos are important. A polymerase with 3′-5′ exonuclease activity may have the ability to remove 3′ modifications that do not provide a substrate for DNA polymerisation. If the rest of the blocking oligo is composed of normal DNA or LNA, then this may allow it to prime synthesis again. Polymerases with 5′-3′ exonuclease activity or strand displacement activity will remove blocking oligos bound downstream of them. A Stoffel fragment polymerase (20) which lack 5′-3′ and 3′-5′ exonuclease activity can be used for this reason in conjunction with DNA or LNA blocking oligos with mismatched 3′-ends for suppression of polymerase priming (19, 21).

3. Methods

3.1. Designing of Blocking Oligos

3.1.1. Different Types of Blocking Principles

A blocking oligo can be designed to overlap with one of the universal primers (called an annealing inhibiting blocking oligo because when bound to the target, it prevents annealing of the universal primer) (Fig. 2a). It can also be positioned between the two universal primers, called an elongation arrest blocking oligo (Fig. 2b). Design an annealing inhibiting blocking oligo if possible as these have proved to be more efficient than elongation arrest blockers (4, 11, 12). However, a conventional oligonucleotide should generally not be more than 25 bases since very long oligos suffer from high T_m and non-specific annealing (22). If it is

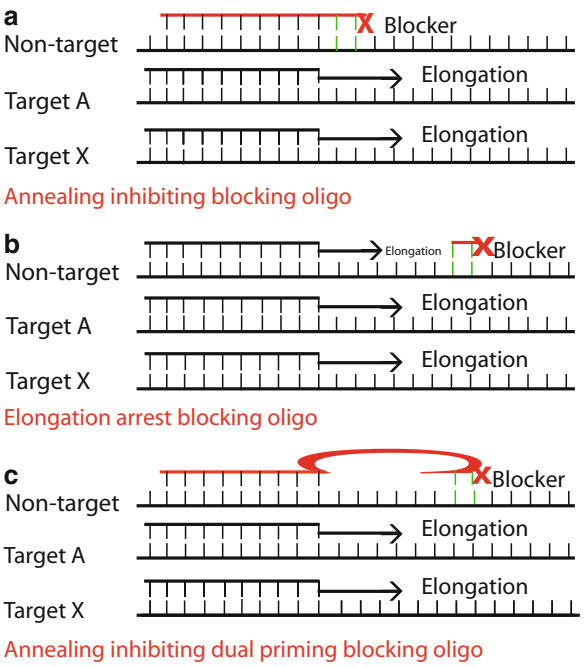


Fig. 2. Modified version from (9). Different mechanisms of blocking oligos. (**a**) Annealing inhibiting blocking oligo. (**b**) Elongation arrest blocking oligo. (**c**) Annealing inhibiting dual priming blocking oligo.

difficult to find a target-specific area in vicinity of the "universal"/target-specific primers used, it is possible to design a dual priming oligo (DPO) ((11, 23), Fig. 2c). This is an oligo where two separate nucleotide strands, one longer 5′ segment and one shorter 3′ segment, are joined by a polydeoxyinosine linker. The longer 5′ segment ensures stable annealing and the shorter 3′ segment makes the DPO very effective in eliminating non-specific priming (23).

3.1.2. Primer Designing

Blocking oligos are designed following the principles of ordinary primer design (e.g. see ref. 22 for general primer design guidelines). In brief, align a representative group of sequences using any preferred multiple alignment software tool (a list is available here: http://bioweb2.pasteur.fr/gensoft/alignment/multiple.html). Then, locate a suitable area in the alignment for placement of the probe. This area should be unique to the species or groups meant to be blocked out. When selecting a blocking oligo binding site, it may be beneficial for the bound oligo to have a T_m slightly higher than that of the primers it is intended to outcompete. However, some care should be taken that it is not so high that it may bind in many regions in the template mix. In order to reduce the risk of "mispriming," runs of three or more Cs or Gs

at the 3′-ends, long runs of a single base, or more than four dinucleotide repeats should be avoided. As with the design of priming oligos, avoid hairpin formation, self-dimerisation, or annealing with any other oligos in the mixture.

3.2. Initial Tests of Blocking Oligo Performance and Optimisation of PCR Conditions

It is important to test that the blocking oligo really is specific to the intended target, as unintentional blocking of other sequences may bias result interpretation, especially in studies that characterise sequence diversity. Check the blocking oligo against sequences in GenBank using BLAST to get a general idea of its specificity. However, given the limited representation of sequences in the database and the uncertainty about the level of sequence identity required for effective blocking at a given annealing temperature, it is important to test the oligo's specificity in vitro with a broad and representative group of DNA samples.

Also, do initial tests of blocking oligo performance to determine the blocker concentration required. The amount of blocker needed for the PCR mixture will depend on the type of blocker chosen (its efficiency), the blocking oligo sequence itself (length and nucleotide composition), and how much non-target DNA there is needed to suppress. As a rule of thumb, ref. 11 shows that a 10:1 ratio of a C3 blocking oligo to universal primers is sufficient for blocking non-target DNA present in 1,000-fold excess to target DNA. If choosing a PNA as the blocking oligo, a much lower ratio, e.g. 1.2:1 (24), can probably be sufficient.

Preferentially, perform the initial tests using artificial DNA mixtures with known and variable concentrations of target DNA to non-target DNA. To create artificial DNA mixtures, first, carry out PCR amplification of target and non-target species using the universal primers only. Clone the PCR products, isolate the plasmids, and linearise them using a restriction enzyme. Linearised plasmid is preferred to PCR product alone as reamplification of PCR products often results in concatemerisation artefacts (artefacts due to self-ligation of small dsDNAs). Quantify the yield and make mixtures of appropriate concentrations (e.g. 100-, 1,000-, and 10,000-fold excess of non-target DNA to target DNA). Use these mixtures as template DNA for new PCRs where blocking oligos are also added to the master mix. Further, annealing temperatures should be initially varied using the gradient feature of the PCR machine to find optimal stringency.

Validation of blocking efficiency can be done by standard gel electrophoresis when using exclusively non-target DNA as a template during testing. Melting curves analysis of the PCR amplicons using SYBR Green fluorescence can be used to differentiate between amplification of target and non-target sequences (see example below). If amplicons differ in size or GC content, methods such as fluorescent fragment analysis on a capillary electrophoresis sequencing platform or denaturing high-performance

liquid chromatography (HPLC) will be much more sensitive and provide semi-quantitative data.

3.3. Methods to Differentiate and Identify the Pool of PCR Products

The products of PCRs generated with blocking oligos from mixed DNA templates generally need to be analysed further to discover the identity, diversity, and sometimes relative quantity of each sequence in the template mix. Methods for achieving this can be classified into those that reveal the primary sequence of the fragments; methods that reveal a secondary characteristic of the fragments such as size or electrophoretic mobility; and methods that employ further reactions to quantify or identify specific sequences. The primary sequences contained in a mixed PCR product have for a long time been characterised by cloning the mixture into plasmids and sequencing plasmids grown from individual bacterial colonies (11). More recently, some next generation sequencing technologies automate this process and when provided with a mixed template sample these can not only avoid the labour-intensive process of cloning and culturing plasmid containing colonies, but also provide far more extensive coverage of diversity present in a sample (3). Secondary characteristics of DNA fragments can be characterised by a wide range of methods including electrophoretic size separation (11, 25); single-strand conformational polymorphism (SSCP, e.g. (25)); denaturing gradient gel electrophoresis (DGGE, e.g. (25, 26)); or high-resolution melting curves associated with SYBR Green fluorescence (27). These methods may not be able to separate or identify all sequence variants, but are very convenient for differentiating between known sequence variants. Quantification and identification of specific amplicons can be achieved using quantitative PCR to detect increases in amplicons with sequence-specific fluorescent "reporter" probes such as TaqMan (28) or similar methods.

3.4. Examples of PCRs Run with and Without Blocking Oligo

The blocking oligo approach has been applied successfully in several studies to allow the detection of minor DNA templates. For example, in an analysis of prey DNA present in seal faecal samples, an annealing inhibiting blocking oligo containing a 3′ C3 spacer was utilised to limit the amplification of predator DNA (3). Real-time SYBR Green PCR with conserved primers showed that addition of the blocking oligo drastically lowered the amount of available template (i.e. seal DNA prevalent in the samples was no longer being amplified and the Ct values shifted to the right; Fig. 3). Melting curve analysis of products confirmed that in PCR without the blocking oligo, primarily seal DNA was amplified; whereas with the blocking probe almost exclusively fish DNA was being amplified (Fig. 3). Sequencing of amplicons from this PCR primer pair/blocking oligo combination found that <3% of the sequences (298/11,248) were derived from predator (3). The same blocking oligo was also used in combination with an alternate

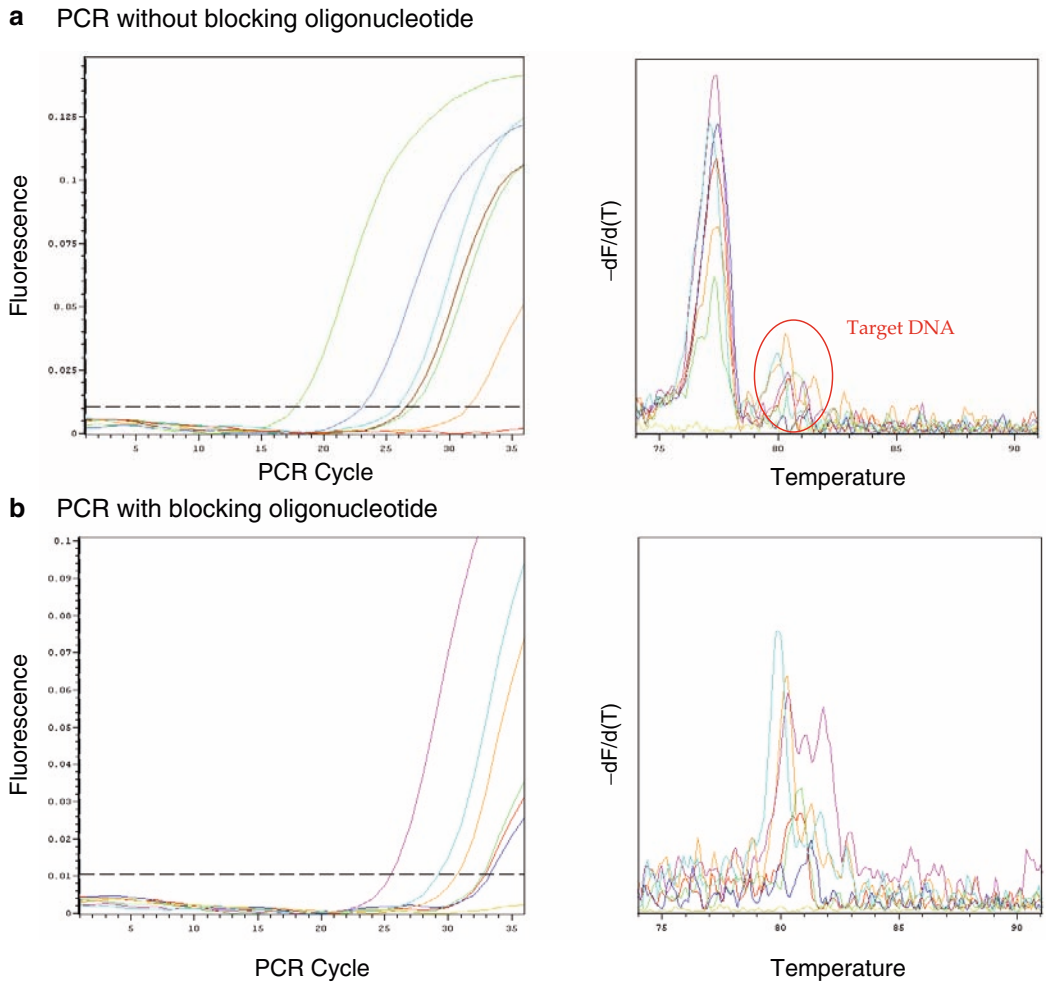


Fig. 3. Real-time PCR amplification profiles and derivative melting curves of resulting PCR products. Both amplifications used the same six DNA extracts from seal faeces as a template and a primer set that amplifies 16S mtDNA from all chordates (Primer Set B from (3)). (**a**) PCR without blocking oligo. (**b**) PCR with the addition of an annealing inhibiting blocking oligo matching the seal sequence (10× concentration relative to PCR primers). With the addition of blocking oligo, the reaction profiles shift to the right due to the reduced amount of template (i.e. seal DNA in the samples is now blocked). The blocking oligo did not have an inhibitory effect on amplification of target DNA (fish genomic DNA; data not shown). Melting profiles show that in the PCR containing the blocking oligo, the dissociation peak matching seal DNA (~77°C) is removed and the peak corresponding to fish DNA (~80–82°C) is expanded.

reverse primer, and post-PCR sequencing showed a higher prevalence of non-target predator DNA (865/3,347, 26% of sequences). The differing effectiveness may be due to change in the ratio of target DNA to non-target DNA in the different PCR amplifications, or a blocking primer may have differing efficiencies depending on the PCR primers used. Regardless, in both cases, the use of blocking oligo meant, prey DNA was the dominant amplicon produced using conserved PCR primers even though DNA from the predator was much more prevalent than prey in the faecal DNA template.

4. Notes

1. Choose a modification which is 100% synthesised (i.e. no oligos missing it) and which is stable (i.e. no degradation or enzymatic removing of the modification after synthesis). Oligos are synthesised backwards in a 3′ to 5′ direction, and choosing a 3′ modification will normally solve the first problem. There are some differences in stability among modifications for making blocking oligos and it appears that the 3′-Spacer C3 CPG is the most stable one.
2. If you design an elongation arrest blocker, a standard oligonucleotide blocking primer should also have a modification on its 5′-end for protection from 5′ exonuclease activity if *Taq* type polymerases are used in amplification.
3. Most standard oligo synthesis reactions are roughly 99% efficient (i.e. each time a nucleotide is added, 99% of the growing oligo chain will receive a nucleotide). This means that after synthesis of a 20-mer, only slightly more than 80% of the products will be full-length oligos, with the rest being incomplete chains. Synthesis of modified oligos outlined above may have reduced synthesis efficiency compared with standard oligos. Since truncated blocking oligos could be less specific than the full-length sequences, purification of full-length oligos after synthesis by HPLC or polyacrylamide gel purification (PAGE) is highly recommended.
4. If you use universal primers targeting a multigene family such as the rDNA operon, note that the different copies within a genome are rarely completely homogeneous (29). The blocking of one version (the most common one and perhaps the only one reported in databases) may cause amplification of pseudogenes or alternative versions (11).

References

1. Bensch, S., Pérez-Tris, J., Waldenström, J., and Hellgren, O. (2004) Linkage between nuclear and mitochondrial DNA sequences in avian malaria parasites: multiple cases of cryptic speciation? *Evolution* **58**, 1617–21.
2. Sidransky, D. (1997) Nucleic acid-based methods for the detection of cancer. *Science* **278**, 1054–58.
3. Deagle, B. E., Kirkwood, R., and Jarman, S. N. (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology* **18**, 2022–38.
4. von Wintzingerode, F., Gobel, U. B., and Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews* **21**, 213–29.
5. Polz, M. F. and Cavanaugh, C. M. (1998) Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology* **64**, 3724–30.
6. King, R. A., Read, D. S., Traugott, M., and Symondson, W. O. C. (2008) Molecular analysis of predation: a review of best practice for DNA-based approaches. *Molecular Ecology* **17**, 947–63.
7. Deagle, B. E., Eveson, J. P., and Jarman, S. N. (2006) Quantification of damage in DNA

recovered from highly degraded samples – a case study on DNA in faeces. *Frontiers in Zoology* **3**, 11.

8. Parsons, B. L. and Heflich, R. H. (1998) Detection of a mouse H-ras codon 61 mutation using a modified allele-specific competitive blocker PCR genotypic selection method. *Mutagenesis* **13**, 581–8.
9. Carlson, C. M., Dupuy, A., Fritz, S., Roberg-Perez, K., Fletcher, C. F., and Largaespada, D. A. (2003) Transposon mutagenesis of the mouse germline. *Genetics* **165**, 243–56.
10. Corless, C. L., Harrell, P., Lacouture, M., Bainbridge, T., Le, C., Gatter, K., White, C., Granter, S., and Heinrich, M. C. (2006) Allele-specific polymerase chain reaction for the imatinib-resistant KIT D816V and D816F mutations in mastocytosis and acute myelogenous leukaemia. *Journal of Molecular Diagnostics* **8**, 604–12.
11. Vestheim, H. and Jarman, S. N. (2008) Blocking primers to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology* **5**, 12.
12. Peano, C., Lesignoli, F., Gulli, M., Corradini, R., Samson, M. C., Marchelli, R., and Marmiroli, N. (2005) Development of a peptide nucleic acid polymerase chain reaction clamping assay for semiquantitative evaluation of genetically modified organism content in food. *Analytical Biochemistry* **344**, 174–82.
13. Karkare, S. and Bhatnagar, D. (2006) Promising nucleic acid analogs and mimics: characteristic features and applications of PNA, LNA, and morpholino. *Applied Microbiology and Biotechnology* **71**, 575–86.
14. Tatsumi, K., Mitani, Y., Watanabe, J., Takakura, H., Hoshit, K., Kawai, Y., Kikuchi, T., Kogo, Y., Oguchi-Katayama, A., Tomaru, Y., Kanamori, H., Baba, M., Ishidao, T., Usui, K., Itoh, M., Cizdziel, P. E., Lezhava, A., Ueda, M., Ichikawa, Y., Endo, I., Togo, S., Shimada, H., and Hayashizaki, Y. (2008) Rapid screening assay for KRAS mutations by the modified smart amplification process. *Journal of Molecular Diagnostics* **10**, 520–6.
15. Ørum H. (2000) PCR clamping. *Current Issues in Molecular Biology*, **1**, 27–30.
16. Troedsson, C., Lee, R. F., Walters, T., Stokes, V., Brinkley, K., Naegele, V., and Frischer, M. E. (2008) Detection and discovery of crustacean parasites in blue crabs (*Callinectes sapidus*) by using 18S rRNA gene-targeted denaturing high-performance liquid chromatography. *Applied and Environmental Microbiology* **74**, 4346–53.
17. Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S. M., Driver, D. A., Berg, R. H., Kim, S. K., Norden, B., and Nielsen, P. E. (1993) PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **365**, 566–8.
18. Ørum, H., Nielsen, P. E., Egholm, M., Berg, R. H., Buchardt, O., and Stanley, C. (1993) Single-base pair mutation analysis by PNA directed PCR clamping. *Nucleic Acids Research* **21**, 5332–6.
19. Dominguez P. L. and Kolodney M. S. (2005) Wild-type blocking polymerase chain reaction for detection of single nucleotide minority mutations from clinical specimens. *Oncogene* **24,** 6830–4.
20. Lawyer, F. C., Stoffel, S., Saiki, R. K., Chang, S. Y., Landre, P. A., Abramson, R. D., and Gelfand, D. H. (1993) High-level expression, purification, and enzymatic characterization of full-length Thermus aquaticus DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. *PCR Methods and Applications* **2**, 275–87.
21. Yu, D., Mukai, M., Liu, Q., and Steinman, C. R. (1997) Specific inhibition of PCR by non-extendable oligonucleotides using a 5' to 3' exonuclease-deficient DNA polymerase. *BioTechniques* **23**, 714–20.
22. Dieffenbach, C. W., Lowe, T. M. J., and Dveksler, G. S. (1993) General concepts for PCR primer design. *PCR-Methods and Applications* **3**, S30–7.
23. Chun, J. Y., Kim, K. J., Hwang, I. T., Kim, Y. J., Lee, D. H., Lee, I. K., and Kim, J. K. (2007) Dual priming oligonucleotide system for the multiplex detection of respiratory viruses and SNP genotyping of CYP2C19 gene. *Nucleic Acids Research* **35**, e40.
24. von Wintzingerode, F., Landt, O., Ehrlich, A., and Gobel, U. B. (2000) Peptide nucleic acid-mediated PCR clamping as a useful supplement in the determination of microbial diversity. *Applied and Environmental Microbiology* **66**, 549–57.
25. Smalla, K., Oros-Sichler, M., Milling, A., Heuer, H., Baumgarte, S., Becker, R., Neuber, G., Kropf, S., Ulrich, A., and Tebbe, C. C. (2007) Bacterial diversity of soils assessed by DGGE, T-RFLP and SSCP fingerprints of PCR-amplified 16S rRNA gene fragments: Do the different methods provide similar results? *Journal of Microbiological Methods* **69**, 470–9.
26. Muyzer, G. (1999) DGGE/TGGE a method for identifying genes from natural ecosystems. *Current Opinion in Microbiology* **2**, 317–22.

27. Robinson, B. S., Monis, P. T., and Dobson, P. J. (2006) Rapid, sensitive, and discriminating identification of *Naegleria* spp. by real-time PCR and melting-curve analysis. *Applied and Environmental Microbiology* **72**, 5857–63.
28. De Jong, W., De Jong, D., and Bodis, M. (2003) A fluorogenic 5' nuclease (TaqMan) assay to assess dosage of a marker tightly linked to red skin color in autotetraploid potato. *Theoretical and Applied Genetics* **107**, 1384–90.
29. Dunshea, G., Barros, N., Wells, R., Gales, N., Hindell, M., and Jarman, S. (2008) Pseudogenes and DNA-based diet analyses: a cautionary tale from a relatively well sampled predator-prey system. *Bulletin of Entomological Research* **98**, 239–48.

Chapter 20

Asymmetric Overlap Extension PCR Method for Site-Directed Mutagenesis

Yue-Hua Xiao and Yan Pei

Abstract

Overlap extension PCR (OE-PCR) has been widely used in site-directed mutagenesis. The original OE-PCR included two rounds of PCRs and required tedious steps to purify the first-round PCR product. By combining asymmetric PCR and overlap extension, a novel asymmetric overlap extension PCR (AOE-PCR) method has been developed. This method consists of two separate asymmetric PCRs of around 30 cycles and a single cycle of annealing and extension after directly mixing the first-round PCR products. AOE-PCR eliminates intermediate purification steps and amplification of wild-type template and requires fewer PCR cycles, and is, therefore, a much simpler and faster and more efficient site-directed mutagenesis method than the original OE-PCR approach.

Key words: Overlap extension PCR, Asymmetric PCR, Asymmetric overlap extension PCR, Site-directed mutagenesis

1. Introduction

Site-directed mutagenesis is one of the most important tools for functional analyses of genes or their regulatory sequences. Various PCR-based protocols have been established for site-directed mutagenesis, among which overlap extension PCR (OE-PCR) and megaprimer PCR are particularly appealing for their simplicity and efficiency (1, 2). Compared with OE-PCR, the megaprimer PCR method is relatively simpler, but generally has a low efficiency (3). Furthermore, the OE-PCR strategies are commonly employed to perform multiple-site mutagenesis (4–6) and are widely used for *in vitro* splicing to generate chimeric genes (7–10).

The original OE-PCR method consists of two rounds of PCRs. For site-directed mutagenesis, two mutant fragments are amplified and purified separately in the first-round PCR, and then

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_20, © Springer Science+Business Media, LLC 2011

annealed and extended as a template in the second-round PCR (11). To simplify the OE-PCR method and simultaneously to increase its efficiency for site-directed mutagenesis, we combined asymmetric PCR and overlap extension to develop a novel asymmetric overlap extension PCR (AOE-PCR) method for site-directed mutagenesis (12). As depicted in Fig. 1, this method consists of two steps, i.e., around 30 cycles of asymmetric PCR and one cycle of annealing and extension. In the first step, two fragments of the mutant DNA are amplified in separate PCRs. In the early cycles, the double-stranded products are exponentially amplified. When the mutant primers of lower concentration are exhausted, the main amplification products in the late cycles would be single strands primed by the flanking primers. Consequently, a lot of single strands with overlapped 3′-ends (mutant primer region, Fig. 1) are generated in the separate amplification reactions. In the second step, the separate reactions are mixed directly without any purification step, and another cycle of annealing and extension is performed with the residual DNA polymerase activities to synthesize adequate target fragments.

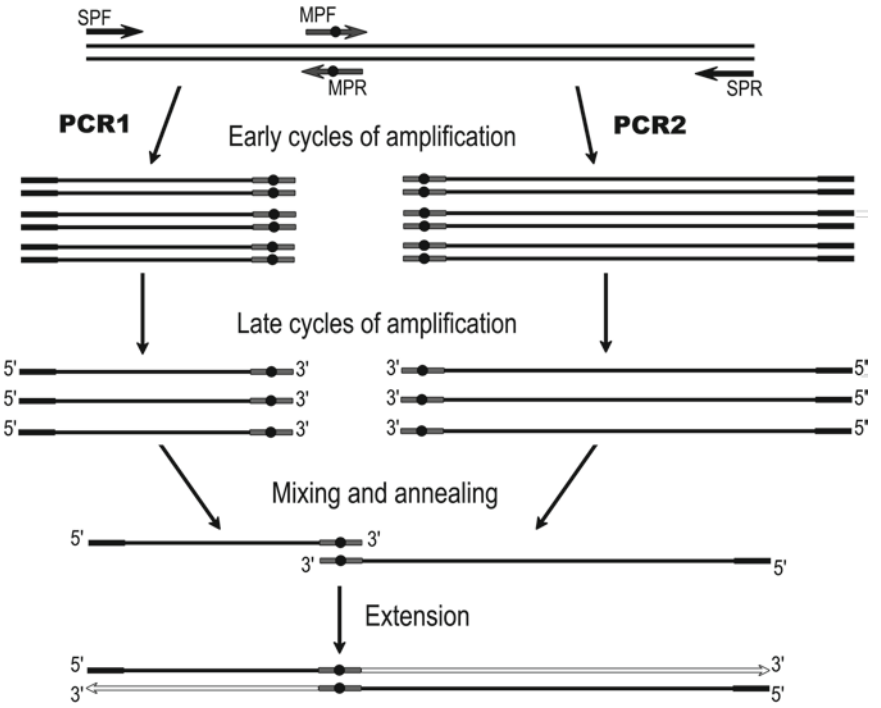


Fig. 1. The schematic representation of the AOE-PCR method. The *black dots* represent the mutant sites. The *open arrows* show the DNA synthesized in the final cycle of extension. The concentrations of mutant primers (forward and reverse mutant primer, MPF and MPR) are lower than those of the flanking primers (forward and reverse specific primer, SPF and SPR). With the mutant primers exhausted in the late cycles of asymmetric PCRs, the major amplification products are the single-stranded DNA primed by flanking primers. After mixing the products of the two asymmetric PCRs, adequate overlap extension products can be synthesized in a single cycle of annealing and extension.

Since the AOE-PCR method consists of only a single cycle of overlap extension without any denaturing step after the mixing of two flanking primers, the amplification of wild-type template is totally excluded. Theoretically, all fragments inserted in AT-cloning vector should be mutated DNA, which guarantees a high mutagenesis efficiency and reduces the cost of selection by sequencing for the site-directed mutagenesis. Secondly, the AOE-PCR method eliminates several tedious steps of the original OE-PCR method, including gel purification of the first-round PCR product and preparation of the second-round PCR. In addition, the second-round PCR in the AOE-PCR method is simplified to a single cycle of annealing and extension, which reduces the total PCR cycles and therefore the polymerase-induced off-target changes in OE-PCR method. Consequently, this new method is much simpler and faster than the original OE-PCR approach. With these advantages, the AOE-PCR method may benefit most of OE-PCR strategies, including those for site-directed mutagenesis as well as *in vitro* splicing (12).

2. Materials

1. TaKaRa LA *Taq* DNA polymerase (TaKaRa, Dalian, China) and 10× LA PCR buffer II (Mg^{2+} plus) with 25 mM Mg^{2+} added by the manufacturer (see Note 1).
2. Oligonucleotide primers including two flanking primers (SPF and SPR) and two mutant primers (MPF and MPR, Fig. 1). All primers should be 20–40 nucleotides in length with a balanced distribution of G and C residues and a low propensity to form stable secondary structure. SPF and SPR reside at the 5′- and 3′-ends of the DNA fragment to be amplified. The mutant bases are introduced by MPF and MPR (see Note 2).
3. dNTP solution containing all four dNTPs, each at 2.5 mM.
4. Gel purification kit (for example, QIAquick Gel Extraction Kit, Qiagen).
5. AT-cloning vectors such as pGEM-T (Promega) and pMD19-T (TaKaRa).

3. Methods

1. Design and synthesize oligonucleotide primers SPF, SPR, MPF, and MPR based on the known sequence of the DNA and intended mutant sequence, as outlined in the protocol of Subheadings 1 and 2 (see Note 2).

2. In a sterile 0.2 ml amplification tube, set up PCR1 as followings (see Note 3).

Template DNA	Approximately 100 ng
10× LA PCR II (Mg^{2+} plus)	5 μl
2.5 mM dNTPs	4.0 μl
5 μM primer SPF	4.0 μl
0.5 μM primer MPR	4.0 μl
TaKaRa LA *Taq*	2 U
H_2O	To 50 μl

3. In a second sterile 0.2 ml amplification tube, set up PCR2 as followings (see Note 3).

Template DNA	Approximately 100 ng
10× LA PCR II (Mg^{2+} plus)	5 μl
2.5 mM dNTPs	4.0 μl
5 μM primer SPR	4.0 μl
0.5 μM primer MPF	4.0 μl
TaKaRa LA *Taq*	2 units
H_2O	To 50 μl

4. Overlay the PCRs with one drop (approximately 30 μl) of light mineral oil. Place the tubes in a thermocycler.
5. Amplify the nucleic acids using the thermocycling conditions listed in Table 1.
6. Set up PCR3 by mixing 25 μl each of PCR1 and PCR2. Overlay the PCRs with one drop (approximately 30 μl) of light mineral oil. Place the tubes in a thermocycler.
7. Join the 5′- and 3′-ends of the mutated DNA by a single cycle of annealing and extension. The times and temperatures are listed in Table 2.
8. Analyze 20 μl of each of the three PCRs by agarose gel electrophoresis and estimate the concentration of amplified target DNAs (Fig. 2, see Note 4).

Table 1
The thermocycling parameters of PCR1 and PCR2

Cycle number	Denaturation	Annealing	Polymerization
1 cycle	5 min at 94°C		
30 cycles	30 s at 94°C	30 s at 55°C	1–2 min at 72°C
Last hold			5 min at 72°C

Table 2
The thermocycling parameters of annealing and extension

Cycle number	Annealing	Polymerization
1 cycle	1 min at 55°C	5 min at 72°C

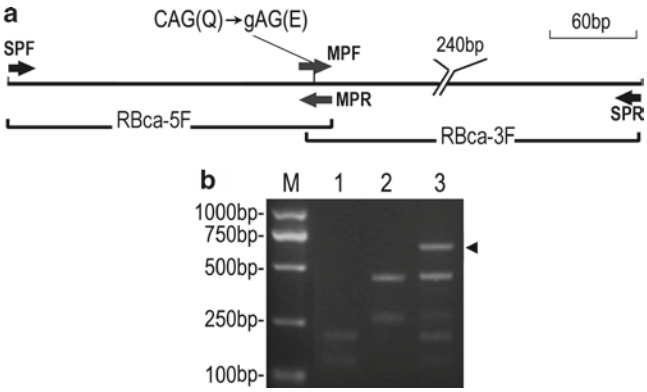


Fig. 2. Introduction of a mutation (Q → E) into a cotton small GTPase gene *GhRacB*. (**a**) The schematic representation of the template and the mutated fragments (5′ and 3′ fragments of GhRacB constitutive active mutant, RBca-5F and RBca-3F). The primers are presented by arrows. The mutant base is introduced by the mutant primers (forward and reverse mutant primers, MPF and MPR). (**b**) Agarose gel analysis of PCR products. *Lanes 1* and *2* contain the asymmetric PCR products (RBca-5F and RBca-3F, respectively); *lane 3* is the overlap extension products. The *arrow head* indicates the target DNA fragments. The positions of DNA marker (*lane M*) are marked on the *left*.

9. Recover the fragment of interest using the gel purification kit (see Note 5), clone the fragment into a TA-cloning vector according to the manufacturer's instructions, and confirm the mutated fragments by sequencing.

4. Notes

1. To avoid the introduction of erroneous bases, use a thermostable DNA polymerase with 3′–5′ exonuclease "proofreading" capacity in overlap extension mutagenesis. TaKaRa LA *Taq* DNA polymerase (TaKaRa) was selected for its high fidelity and the ability to catalyze the nontemplate addition of adenine residues, which facilitated direct TA-cloning of PCR products.
2. Besides general considerations regarding the design of PCR primers, the mutant primers should be 2- to 6-nt longer to compensate the effects of mutant bases on annealing temperature,

and the mutant bases should reside in the middle of the primer to ensure efficient annealing to the wild-type template.

3. Generally, the plasmid containing the wild-type DNA is used as templates in PCR1 and PCR2. After treating at 95°C for 10 min, 1–2 μl of *Escherichia coli* cultures containing the plasmid could also be employed as a template.
4. Analyze products of the three PCRs in parallel on a single agarose gel, so that the target fragment can be easily identified. The single-stranded products of asymmetric PCR appear as bands migrating faster through agarose gels than the corresponding double-stranded fragment.
5. Repeat AOE-PCR (steps 2–8) with another pair of mutant primers using the gel-purified or nonpurified AOE-PCR products as a template if you need to introduce a second mutant (12).

References

1. Ling, M. M. and Robinson, B. H. (1997) Approaches to DNA mutagenesis: an overview. *Anal. Biochem.* **254**, 157–78.
2. Sambrook, J. and Russell, D. W. (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
3. Andag, R. and Schutz, E. (2001) General method for site-directed mutagenesis. *BioTechniques* **30**, 486–8.
4. Ge, L. and Rudolph, P. (1997) Simultaneous introduction of multiple mutations using overlap extension PCR. *BioTechniques* **22**, 28–30.
5. An, Y., Ji, J., Wu, W., Lv, A., Huang, R., and Wei, Y. (2005) A rapid and efficient method for multiple-site mutagenesis with a modified overlap extension PCR. *Appl. Microbiol. Biotechnol.* **68**, 774–8.
6. Chang, S. W., Shieh, C. J., Lee, G. C., and Shaw, J. F. (2005) Multiple mutagenesis of the *Candida rugosa* LIP1 gene and optimum production of recombinant LIP1 expressed in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* **67**, 215–24.
7. Warrens, A. N., Jones, M. D., and Lechetler, R. I. (1997) Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. *Gene* **186**, 29–35.
8. Mehta, R. K. and Singh, J. (1999) Bridge-overlap-extension PCR method for constructing chimeric genes. *BioTechniques* **26**, 1082–6.
9. Kuwayama, H., Obara, S., Morio, T., Katoh, M., Urushihara, H., and Tanaka, Y. (2002) PCR-mediated generation of a gene disruption construct without the use of DNA ligase and plasmid vectors. *Nucleic Acids Res.* **30**, e2.
10. Shevchuk, N. A., Bryksin, A. V., Nusinovich, Y. A., Cabello, F. C., Sutherland M., and Ladisch, S. (2004) Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. *Nucleic Acids Res.* **32**, e19.
11. Ho S. N., Hunt, H. D., Horton, R. M., Pullen J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–9.
12. Xiao, Y. H., Yin, M. H., Hou, L., Luo, M., and Pei, Y. (2007) Asymmetric overlap extension PCR method bypassing intermediate purification and the amplification of wild-type template in site-directed mutagenesis. *Biotechnol. Lett.* **29**, 925–30.

Chapter 21

Ribosome Display: A Technology for Selecting and Evolving Proteins from Large Libraries

Birgit Dreier and Andreas Plückthun

Abstract

The selection and concomitant affinity maturation of proteins to bind to user-defined target molecules have become a key technology in biochemical research, diagnostics, and therapy. One of the most potent selection technologies for such applications is ribosome display. It works entirely *in vitro*, and this has two important consequences. First, since no transformation of any cells is required, libraries with much greater diversity can be handled than with most other techniques. Second, since a library does not have to be cloned and transformed, it is very convenient to introduce random errors in the library by PCR-based methods and select improved binders. Thus, a true directed evolution, an iteration between randomization and selection over several generations, can be conveniently carried out, e.g., for affinity maturation. Ribosome display has been used successfully for the selection of antibody fragments and other binding proteins, such as *D*esigned *A*nkyrin *R*epeat *P*roteins (DARPins).

Key words: Ribosome display, *In vitro* selection, *In vitro* translation, Designed ankyrin repeat proteins, Affinity maturation

1. Introduction

In order to select and evolve proteins or peptides from a library to bind to any chosen target of interest, different selection strategies can be applied. All technologies have in common that the phenotype (peptide or protein scaffold) is physically linked to the genetic information (DNA or mRNA). We term "*in vivo*" those technologies that require transformation of cells with a library, as needed, for example, in phage display (1). In contrast, "*in vitro*" technologies do not require any transformation of cells with the library. Examples are ribosome display (2, 3) or mRNA display (4, 5). In each case, selection of highly specific binders is performed over multiple rounds of selection, starting from a library of peptides or proteins with a natural or designed scaffold (Fig. 1).

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_21, © Springer Science+Business Media, LLC 2011

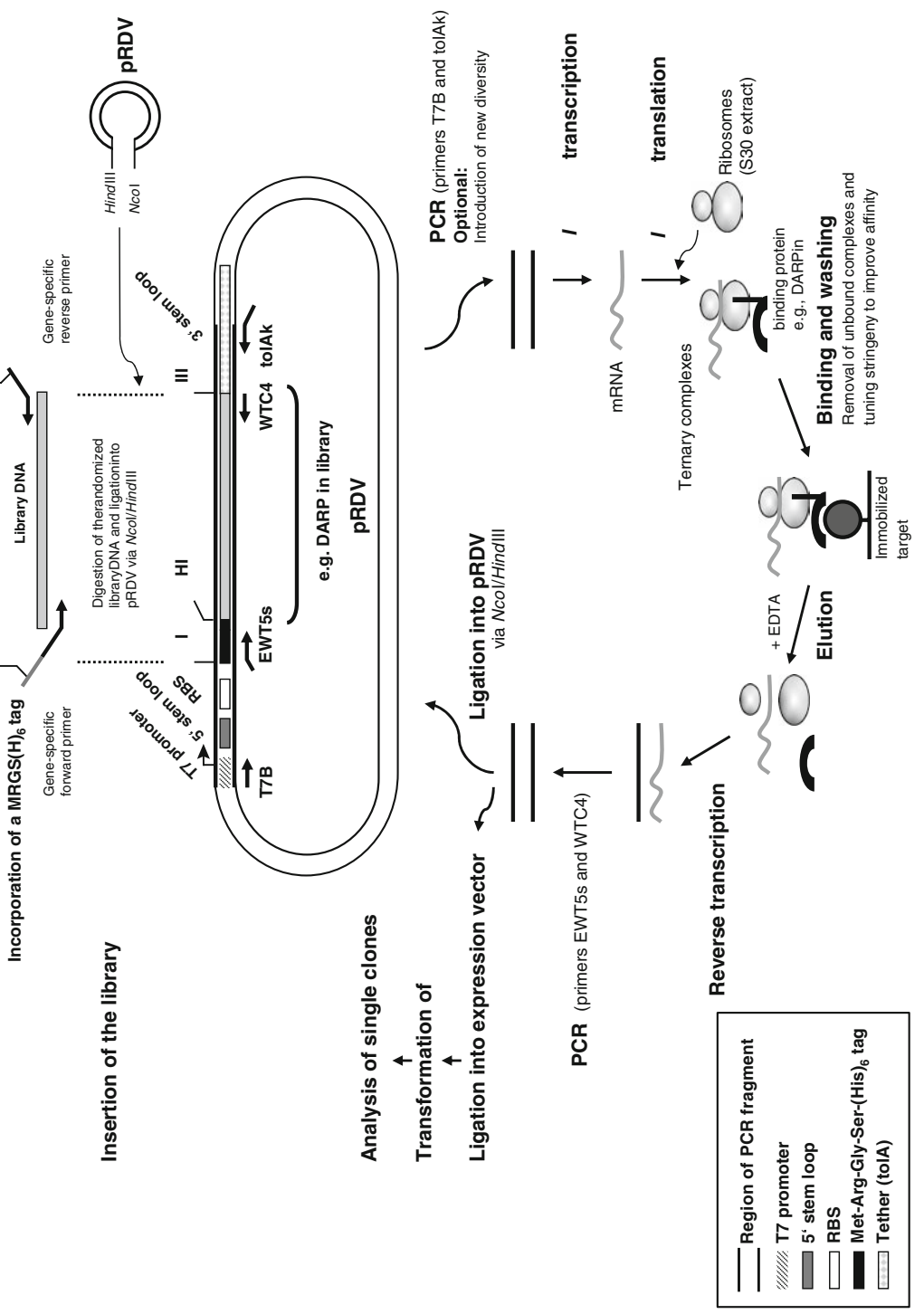

Incorporation of a MRGS(H)6 tag
Gene-specific forward primer
Library DNA
Gene-specific reverse primer
Insertion of the library
Digestion of therandomized libraryDNA and ligationinto pRDV via NcoI/HindIII
pRDV
HindIII
NcoI
3' stem loop
T7 promoter
5' stem loop
RBS
EWT5s
T7B
e.g. DARP in library
pRDV
WTC4
tolAk
Analysis of single clones
Transformation of
Ligation into expression vector
Ligation into pRDV
via NcoI/HindIII
PCR (primers T7B and tolAk)
Optional:
Introduction of new diversity
transcription
translation
mRNA
Ribosomes
(S30 extract)
Ternary complexes
binding protein
e.g., DARPin
Binding and washing
Removal of unbound complexes and
tuning stringency to improve affinity
Immobilized
target
Elution
+ EDTA
PCR (primers EWT5s and WTC4)
Reverse transcription
Region of PCR fragment
T7 promoter
5' stem loop
RBS
Met-Arg-Gly-Ser-(His)6 tag
Tether (tolA)

Fig. 1. Scheme of the ribosome display cycle, illustrated for selection of high-affinity DARPins. In ribosome display all steps of the selection are performed *in vitro*. The cycle begins with a DNA library (*top*) in the form of a PCR fragment encoding a library of the protein of interest. This cassette is ligated into a vector *in vitro*, which provides a promoter and ribosome binding site. The ORF of interest (*light gray*) is fused to an additional protein region (the "spacer" or "tether," *checkered white*). This tether or spacer, here used as an unstructured region from the *E. coli* TolA protein, has the sole function of allowing the protein domain of interest to emerge from the ribosomal tunnel. A PCR is then carried out from the promoter to the middle of the tether. Importantly, the PCR fragment does not encode a stop codon at the end. Each member of the library pool is then transcribed from double-stranded DNA into mRNA and is subsequently translated by the ribosomes present in the S30 extract, leading to ternary complexes consisting of ribosomes, mRNA, and the DARPin encoded by that particular mRNA. Since there is no stop codon on the mRNA, the protein is not released from the ribosome. It is believed to be still covalently attached to the tRNA within the ribosome, with the tether in the tunnel, and the domain of interest outside and already folded. Selection can be achieved by binding the protein–ribosome–mRNA complexes to the desired immobilized target, followed by removal of unbound or nonspecifically bound protein by stringent washing. Affinity can be increased by addition of an excess of nonlabeled target (off-rate selection) (see Subheading 3.5.4). Particular selectivity in binding can be achieved by adding an unwanted target as a competitor. Selection for other properties, such as stability, requires other selection pressures at this step (see Note 25). Binders can be easily recovered by destruction of the protein–ribosome–mRNA complex using EDTA and recovery of the genetic information of the binders by RT-PCR using the inner primers WTC4 (annealing to the sequence encoding the C-terminus of the DARPin sequence which can be replaced by a primer specific for other library folds) and EWT5s [pRDV-specific primer overlapping with the ribosome binding site (RBS) and beginning of the Met-Arg-Gly-Ser-$(His)_6$ tag]. The inner primer set is used to amplify the selected clones, which often is not possible with the outer primer set due to incomplete synthesis or degradation of the mRNA. For further selection rounds, the PCR product pool is subcloned into pRDV via the restriction endonucleases *Nco*I and *Hin*dIII, followed by a second PCR with the outer primers T7B and tolAk. T7B introduces the T7 promoter sequence and part of the stabilizing 5′ stem loop, sequences that are part of the pRDV vector. The tolAk primer binds in the sequence of the tolA spacer region and introduces a stabilizing 3' stem loop. If further diversity is required an error-prone PCR can be included at this step. The amplified PCR product then serves as template for *in vitro* transcription, initiating the next round of selection. At the end of the selection rounds (typically, 2–5), the resulting PCR product pool can be directly subcloned via the restriction endonucleases *Bam*HI and *Hin*dIII into an expression vector in order to screen for binders.

Ribosome display has some major advantages compared to *in vivo* selection strategies. First, the library size is not restricted to limitations in transformation efficiency, which for *Escherichia coli* usually is 10^9–10^{10} per microgram of DNA and significantly lower when a ligation mixture is used. In ribosome display, library size is only limited by the number of ribosomes present in the *in vitro* translation and can be as high as 10^{12}–10^{14}. Second, while in the living cell protein biosynthesis and folding occur in a given environment, the environment using an *in vitro* selection system can be manipulated and optimized for expression, folding, and stability of the library members. This can be exploited for stability selection (6, 7). Third, the diversity of the library members can be easily manipulated at any selection step by introduction of additional mutations using DNA shuffling (8) and/or error-prone PCR (9). This is perhaps the factor of greatest practical utility. In other technologies, after each randomization step, a new library needs to be ligated and transformed. In ribosome display, an additional randomization step merely alters the method of library amplification. Therefore, ribosome display is particularly suited for directed evolution projects over many generations.

Ribosome display selections have been first exploited for peptides (10), but the true advantage of directed evolution by using error-prone PCR methods was only borne out with proteins, and antibody scFv fragments (3, 7, 11, 12) were the first protein used. By using very stringent selections for affinity, antibody scFv fragments have been evolved to affinities as high as 5 pM or even 1 pM (12, 13). By choosing an appropriate selection pressure, properties other than affinity could be optimized, such as, for example, stability (7).

Over the last years also new protein scaffolds, with more desirable biophysical properties than antibody fragments, have been investigated. Among the most promising scaffolds are the *D*esigned *A*nkyrin *R*epeat *P*roteins (DARPins), which are devoid of disulfide bonds, highly soluble, and highly stable, and therefore achieve high expression levels in *E. coli* (14). They also fold well in the *in vitro* translation inherent in ribosome display. Using ribosome display DARPins have been evolved to bind various targets with affinities all the way down to the picomolar range (15–20). In general, probably because of their robust *in vitro* folding, DARPins are enriched over fewer rounds than antibody fragments.

Taken together, ribosome display is an ideal tool to select and evolve proteins with predefined binding properties from large libraries.

2. Materials

2.1. General

1. 96-well Maxisorp plates (Nunc, No. 442404) or strips.
2. Adhesive plate sealers (Thermo Scientific, No. AB-0580).

3. Sterile, RNase-free ART filter tips (Molecular Bio Products).
4. Sterile, RNase-free HydroLogix 1.5 and 2.0-ml tubes (Molecular Bio Products, No. 3448 or No. 3434).
5. Roche high pure RNA isolation kit (Roche, No. 11828665001).
6. illustra MicroSpin™ G-50 Columns (GE Healthcare, No. 27-5330-01).
7. NucleoSpin® Extract II DNA purification kit (Macherey-Nagel, No. 740609.50).

2.2. Reagents for Selection (See Note 1)

1. Tris-buffered saline (TBS): 50 mM Tris, 150 mM NaCl; adjust pH to 7.4 with HCl at 4°C; filter through 0.22 μm.
2. TBST: TBS containing 0.05% Tween-20.
3. Stock solutions for wash buffer (WB) and elution buffer (EB): 2 M Tris-acetate; adjust pH to 7.5 at 4°C with acetic acid, 5 M NaCl, 2 M magnesium acetate, 250 mM EDTA; adjust pH to 8.0 by NaOH addition. Sterile filter all solutions. For alternative buffer composition see Note 2.
4. WB/Tween-20 (WBT): 50 mM Tris-acetate pH 7.5, 150 mM NaCl, 50 mM magnesium acetate, 0.05% Tween-20; adjust pH to 7.5 with acetic acid at 4°C; filter through 0.22 μm.
5. EB: 50 mM Tris-acetate pH 7.5, 150 mM NaCl, 25 mM EDTA; adjust pH to 7.5 with HCl at 4°C; filter through 0.22 μm.
6. *Saccharomyces cerevisiae* RNA (BioChemica, No. 83847): dissolve to 25 μg/μl in H_2O, aliquot and store at −20°C.
7. 10% BSA in H_2O: filter through 0.22 μm and store at −20°C.
8. Neutravidin and/or streptavidin (Pierce, No. 31000 or No. 21125): 1.2 mg/ml (20 μM) in TBS and store at −20°C.
9. Streptavidin-coated magnetic beads (MyOne T1; Invitrogen, No. 65602).
10. Reagents for biotinylation of the target: either for chemical biotinylation a NHS-biotin reagent [e.g., from Pierce EZ-link™ SulfoNHS-LC-biotin (No. 21335)] or for enzymatic biotinylation of an AviTag using the *E. coli* biotinylation enzyme BirA (21) (reagents from Avidity).

2.3. Reagents for mRNA Cleanup After In Vitro Transcription

1. 6 M LiCl; filter through 0.22 μm pores.
2. 3 M sodium acetate; filter through 0.22 μm pores.
3. 70% EtOH diluted with H_2O and 100% EtOH; filter through 0.22 μm pores.
4. illustra MicroSpin™ G-50 Columns (GE Healthcare, No. 27-5330-01).
5. DNAseI (10 U/μl; Roche, No. 04716728001).

2.4. Reagents for Reverse Transcription, PCR and Cloning

1. Primer dissolved to 100 µM in H_2O; aliquot and store at –20°C.
   EWT5s: 5′-TTCCTCCATGGGTATGAGAGGATCG-3′
   WTC4: 5′-TTTGGGAAGCTTTTGCAGGATTTCAGC-3′
   T7B: 5′-ATACGAAATTAATACGACTCACTATAGGGAGA CCACAACGG-3′
   tolAk: 5′-CCGCACACCAGTAAGGTGTGCGGTTTCAG-TTGCCGCTTTCTTTCT-3′
2. AffinityScript™ Multiple Temperature Reverse Transcriptase (50 U/µl; Stratagene, No. 600107) and 10× buffer; see Note 3.
3. 100 mM DTT in H_2O; aliquot and store at –20°C.
4. RNasin® Ribonuclease Inhibitor (20–40 U/µl; Promega, No. N2115).
5. $Vent_R$® DNA Polymerase (2 U/µl; New England Biolabs, No. M0254S) and 10× Thermopol buffer; see Note 4.
6. Platinum® Taq DNA Polymerase (5 U/µl; Invitrogen, No. 10966083) and 10× polymerase buffer.
7. dNTPs: 5 mM each (Eurogentec, No. NU-0010-10); aliquot and store at –20°C.
8. Nucleotide analogs dPTP and 8-oxo-dGTP (Jena Biosciences) at a concentration of 100 µm.
9. Dimethyl sulfoxide (DMSO; Fluka, 41640).
10. Restriction endonucleases: *Bam*HI (20 U/µl; No. R0136L), *Hin*dIII (20 U/µl; No. R0104L), *Nco*I (10 U/µl; No. R0193L), and 10× buffer all from New England Biolabs.
11. T4 DNA ligase (5 U/µl; Fermentas, No. EL0014) and 10× ligase buffer.
12. Ribosome display vector pRDV (GenBank accession code AY327136; please note the revised sequence) (16).

2.5. Reagents for In Vitro Transcription

1. T7 RNA polymerase (20 U/µl; Fermentas, No. EP0111); see Note 5.
2. RNasin® Ribonuclease Inhibitor (20–40 U/µl; Promega, No. N2115).
3. 100 mM DTT in H_2O; aliquot and store at –20°C.
4. T7 RNA polymerase buffer (5×): 1 M HEPES, 150 mM magnesium acetate, 10 mM spermidine, 200 mM DTT; adjust pH to 7.6 with KOH; aliquot and store at –20°C.
5. 50 mM NTP mix: 50 mM adenosine 5′-triphosphate (ATP; Sigma-Aldrich, No. A2383), 50 mM uridine 5′-triphosphate (UTP; Sigma-Aldrich, No. U6625), 50 mM guanosine 5′-triphosphate (GTP; Sigma-Aldrich, No. G8877), 50 mM cytidine 5′-triphosphate (CTP; Sigma-Aldrich, No. C1506) in H_2O; aliquot and store at –20°C.

2.6. Reagents for In Vitro Translation

1. Protein disulfide isomerase (PDI; Sigma-Aldrich, No. P3818): 22 µM in H_2O; aliquot and store at −80°C.
2. Heparin (Sigma-Aldrich, No. H4784) stock solution: 200 mg/ml heparin in H_2O (do not filter); aliquot and store at −20°C.
3. Methionine (Sigma-Aldrich, No. M9625): 200 mM l-methionine in H_2O (do not filter); aliquot and store at −20°C.
4. STOP mix: 1 ml WBT buffer/0.5% BSA plus 12.5 µl heparin stock solution (see above).

2.6.1. S30 Extract

1. *E. coli* strain MRE600 (ATCC 29417) (22) lacking ribonuclease I activity.
2. Incomplete rich medium: 5.6 g KH_2PO_4, 28.9 g K_2HPO_4, 10 g yeast extract, 15 mg thiamine for 1 l medium. Autoclave and add 50 ml 40 % glucose (w/v) and 10 ml 0.1 M magnesium acetate, both sterile filtered.
3. S30 buffer: 10 mM Tris-acetate (pH 7.5 at 4°C), 14 mM magnesium acetate, 60 mM potassium acetate. Chill to 4°C before use.
4. Preincubation mix (must be prepared directly before use): 3.75 ml 2 mM Tris-acetate (pH 7.5 at 4°C), 71 µl 3 M magnesium acetate, 75 µl amino acid mix (10 mM of each of the 20 amino acids; Fluka, No. LAA21), 300 µl 0.2 M ATP, 50 U pyruvate kinase (Fluka, No. 83328), 0.2 g phosphoenolpyruvate trisodium salt (Fluka, No. 79435); add to 10 ml H_2O.

2.6.2. PremixZ

1. Set up premixA (the final concentration will be fivefold lower in the final volume of the *in vitro* translation reaction; see Subheading 3.4): 250 mM Tris-acetate (from a 2 M stock solution, pH 7.5 at 4°C), 18 µM anti-ssrA oligonucleotide (5′-TTAAGCTGCTAAAGCGTAGTTTTCGTCGTTTGCGACTA-3′) from a 200 µM stock solution, 1.75 mM of each amino acid except for methionine, 10 mM ATP from an 1 M stock solution, 2.5 mM GTP from a 0.2 M stock solution, 5 mM cAMP (Sigma-Aldrich, No. A6885) from 0.4 M stock solution, 150 mM acetyl phosphate (Sigma-Aldrich, No. A0262) from 2 M stock solution, 2.5 mg/ml *E. coli* tRNA from strain MRE600 (Roche, No. 10109541001) from a 25 mg/ml stock solution, 0.1 mg/ml folinic acid (Sigma-Aldrich, No. 47612) from 10 mg/ml stock solution.
2. Set up an *in vitro* translation reaction (see Subheading 3.4), and use the above premixA but titrate the optimal concentration of the following components for the final premixZ composition to achieve optimal performance of each newly generated S30 extract. Optimize the final concentrations in the order shown:

Magnesium acetate (MgAc) usually in the range of 7–15 mM from a 0.2 M stock solution, potassium glutamate (KGlu) usually in the range of 180–220 mM from a 2 M stock solution, and PEG-8000 usually in the range of 5–15% (w/v) from a 40% stock solution. Adjust the premixA with the optimal composition of MgAc, KGlu, and PEG to obtain the premixZ. Aliquot the premixZ and flash-freeze in liquid nitrogen. Long-time storage should be at –80°C, but the premixZ is stable for several months at –20°C and can be frozen several times. If not noted otherwise, reagents were purchased from Sigma-Aldrich.

2.6.3. β-Lactamase Assay

Used to test the activity of the S30 extract and optimization of the premixZ.

1. Prepare β-lactamase mRNA from the pRDV template DNA encoding the double Cys → Ala mutant of β-lactamase (23) using PCR with the T7B and tolAk primers (Fig. 1) (see Subheading 3.2), followed by *in vitro* transcription and purification of mRNA (Protocol 3.2 and 3.3).
2. Set up *in vitro* translation reactions containing 2 µg RNA, 0.5 µl 200 mM methionine, 10 µl S30 extract, 8.2 µl premixZ and add to 22 µl H_2O. For optimization of the activity of the S30 extract use premixA and adjust the concentration of magnesium acetate, potassium acetate, and PEG-8000.
3. Incubate at 37°C for 10 min.
4. Add 88 µl STOP mix
5. Use 5 µl of stopped *in vitro* translation for the activity assay with the chromogenic substrate nitrocefin (Glaxo Research No. 87/312, obtained from Oxoid No. SR0112) (24).
6. Dilute nitrocefin 1:20 in β-lactamase buffer (100 mM sodium phosphate buffer, pH 7.0) from a stock solution (1 mg nitrocefin dissolved in 500 µl DMSO and stored at –20°C). For one reaction use 20 µl diluted nitrocefin together with 5 µl translation plus 175 µl β-lactamase buffer in a 200 µl reaction.
7. Measure $OD_{486\,nm}$ immediately. Follow the kinetics for approximately 12 min, measuring at least once every minute.

2.7. Reagents for DARPin Expression and Binding Analysis of Single Clones (25)

1. *E. coli* strain XL-1 blue (Stratagene, No. 200268).
2. Expression plasmid pDST67 (20, 25), a derivative of pQE30 (QIAGEN).
3. 2×TY media: 5 g NaCl, 16 g tryptone, 10 g yeast extract per liter. Adjust pH to 7.2 with NaOH.
4. 96-well deep well plates (ABgene, No. AB-0661).
5. Tris-buffered saline (TBS): 50 mM Tris, 150 mM NaCl; adjust pH to 7.4 with HCl.
6. TBST: TBS containing 0.05% Tween-20.

7. 10% BSA in H_2O.
8. Mouse-anti-RGS$(His)_4$ antibody (QIAGEN, No. 34650).
9. Goat-anti-mouse IgG coupled to alkaline phosphatase (Sigma-Aldrich, No. A3562).
10. pNPP substrate (*p*-nitrophenyl phosphate disodium salt; Fluka, No. 71768): stock 1 M in pNPP buffer (50 mM $NaHCO_3$, 50 mM $MgCl_2$); aliquot and store at −20°C.
11. B-PER II detergent solution (Pierce, No. 78260).

3. Methods

3.1. Insertion of the Library

The ribosome display vector pRDV is used to ligate the library of interest using gene-specific primers and insertion via the restriction endonuclease sites *Bam*HI and *Hin*dIII as indicated in Fig. 1 (16, 26). The general elements that need to be present in a ribosome display vector are the T7 RNA polymerase promoter sequence to initiate efficient transcription and a RBS for docking of the ribosome to initiate translation. The PCR fragment (between the primers T7B and tolAk, Fig. 1) that serves as the template for transcription ends without a stop codon in the ORF. At both the 5′ and 3′ ends of the mRNA, stabilizing stem loops are incorporated to protect the mRNA from exonuclease degradation (2, 27). The absence of a stop codon in the resulting mRNA prevents termination of translation. The fact that the library is fused in frame to a spacer (or tether) sequence (e.g., derived from the *E. coli tolA* gene) allows the nascent protein chain to exit the ribosome and fold outside of the ribosome. The original pRDV contains an N-terminal FLAG tag instead of an N-terminal Met-Arg-Gly-Ser-$(His)_6$ tag as shown here for the case of the DARPin libraries (16). Both tag variants lead to good initiation of *in vitro* translation and yield ternary complexes in good yields.

3.2. Transcription of PCR Products

1. To obtain a length-defined fragment of DNA as template for *in vitro* transcription, use the outer primers T7B and tolAk in the following PCR reaction to introduce the T7 RNA polymerase promoter sequence, RBS, the stabilizing 5′ and 3′ stem loops and the tolA spacer sequence:

   5.0 µl 10× Thermopol buffer

   2.0 µl dNTPs (final concentration 200 µM each)

   2.0 µl DMSO (final concentration 5%)

   0.5 µl T7B primer (final concentration 1 µM)

   0.5 µl tolAk primer (final concentration 1 µM)

5.0 µl library DNA [either of the initial library or of the amplified DNA after selection which has been ligated to pRDV (see Note 6)]

0.5 µl Vent DNA polymerase (2 U/µl)

Add to 50 µl H_2O.

2. Perform a hot start to increase specificity and use the following cycling parameters (see Note 7): 3 min at 95°C, 25 cycles: 30 s at 95°C, 30 s at 55°C, 45 s at 72°C, final extension 5 min at 95 °C.
3. Verify the product on an agarose gel.
4. For *in vitro* transcription set up the following reaction on ice:

   20 µl 5× T7 polymerase buffer

   14 µl NTPs (final concentration 7 mM each)

   4 µl T7 RNA polymerase (20 U/µl)

   2 µl RNasin (40 U/µl)

   22.5 µl PCR product without further purification

   Add to 100 µl with H_2O.
5. Incubate the transcription for 2–3 h at 37°C (see Note 8).

3.3. Cleanup of Template mRNA for In Vitro Translation

1. In order to remove all impurities from the reaction, the RNA needs to be purified. This can be performed in two ways:

   Conventional protocol

   (a) A LiCl precipitation can be performed to purify the RNA product. For this purpose, add 100 µl ice-cold H_2O and 200 µl ice-cold 6 M LiCl to the 100 µl translation reaction and vortex.

   (b) Incubate on ice for 30 min, then centrifuge at 20,000 × *g* at 4°C for 30 min.

   (c) Discard the supernatant and wash the pellet with 500 µl ice-cold 70% EtOH ensuring that the pellet is not disturbed.

   (d) Remove supernatant and dry pellet in a Speedvac apparatus.

   (e) Completely dissolve the pellet in 200 µl ice-cold H_2O and centrifuge at 20,000 × *g* at 4°C for 5 min to remove remaining precipitates.

   (f) Transfer 180 µl supernatant to a new tube without disturbing the pellet. Add 20 µl 3 M NaOAc and 500 µl ice-cold 100% EtOH, vortex.

   (g) Incubate at –20°C for at least 30 min. Vortex and centrifuge at 20,000 × *g* at 4°C for 30 min and discard the supernatant.

   (h) Wash the pellet with 500 µl ice-cold 70% EtOH, dry the pellet in a Speedvac apparatus, and resuspend the pellet in 30 µl H_2O.

Alternative protocol

(a) For purification of the RNA, small gel filtration columns (e.g., illustra MicroSpin™ G-50 Columns) can be used.

(b) Vortex the column to resuspend the material and break off bottom of the column.

(c) Place the column into a 1.5-ml tube and spin down at 735 × *g* for 1 min to pack the column material.

(d) Place the column into a collection tube, apply 50 µl sample from the transcription reaction, and centrifuge at 735 × *g* for 1 min.

(e) Optional: DNAse I treatment before loading the column (see Note 9): Take 43 µl of the transcription reaction and add 2 µl of DNAse I solution (10 U/µl) plus 5 µl 10× dilution buffer supplied with the enzyme. Incubate for 10–15 min at room temperature, and then apply the sample to the column.

2. Aliquot RNA and immediately freeze in liquid nitrogen. Store at –80°C.

3. Determine the RNA concentration of a 1:100 dilution by $OD_{260\,nm}$. If the transcription worked well, a yield of 3–8 µg/µl for RNA after LiCl/EtOH precipitation (total yield from a 100 µl reaction: 90–240 µg) or 1–3 µg/µl from the illustra MicroSpin™ G-50 Columns (total yield from a 50 µl reaction: 50–150 µg) should be obtained.

3.4. In Vitro Translation

1. For one *in vitro* translation reaction, set up the following mix on ice:

   2.0 µl 200 mM methionine

   41 µl premixZ with optimized composition

   x µl *in vitro* transcribed RNA (total 10 µg, volume follows from RNA concentration; see Note 10)

   50 µl S30 extract and add to 110 µl with H_2O (for preparation of the S30 extract, see Subheading 3.12)

   Add 0.625 µl PDI if your library scaffold requires the formation of disulfide bonds.

2. Mix carefully by pipetting up and down and incubate the reaction at 37°C for 10 min, the time found optimal for DARPins. The incubation time and temperature must be optimized for each library based on different constructs.

3. Stop the reaction by addition of 440 µl ice-cold STOP mix.

4. Mix by pipetting up and down and centrifuge at 20,000 × *g* at 4°C for 5 min. Transfer 500 µl supernatant to a fresh tube and use 100 µl per well when performing selection in plates or 250 µl per tube when performing selections in solution for either the target-containing or control reaction (see Subheading 3.5).

3.5. Selection (See Note 11)

3.5.1. Target Protein Preparation

Express and purify the target by methods of your choice. To immobilize the target for capturing the ternary complexes it is recommended to biotinylate the target. This is the method of immobilization found to be most robust by far to stringent washing, including washing with detergents. The advantage of immobilizing biotinylated targets is that it is very general, and it works equally well for proteins, peptides, oligonucleotides, and small molecules. Furthermore, by avoiding any direct binding to plastic surfaces, the structure of the target is maintained. Finally, the nonbiotinylated version of the target is a convenient competitor in off-rate selections and in the specificity screening of single clones. Biotinylation can be achieved in two ways (see Note 12):

1. Fuse the target to an AviTag and biotinylate it *in vivo* or *in vitro* using the *E. coli* biotinylation enzyme BirA (21) following the guidelines posted on the Avidity webpage (http://www.avidity.com).
2. Alternatively, biotinylate surface lysine amino acid residues using NHS-biotin reagents from Pierce following the manufacturer's instructions.

3.5.2. Selection in Plates

1. Coat wells of a 96-well Maxisorp plate with 100 µl of a 66 nM neutravidin or streptavidin solution in TBS and close with an adhesive plate sealer (see Notes 13 and 14). Store overnight at 4°C or for 1 h at room temperature. Invert the plate and shake out the solution, dry on paper towels, and wash the wells three times with 300 µl TBS.
2. Block the wells with 300 µl 0.5% BSA in TBST per well, seal and incubate on an orbital shaker for 1 h at room temperature. Shake out blocking solution and dry on paper towels.
3. Immobilize 100 µl biotinylated target at a concentration of 100–200 nM (can be decreased in later rounds) in TBST/0.5% BSA and TBST/0.5% BSA only for control wells. Seal and incubate on an orbital shaker at 4°C for 1 h. Wash the plate three times with 300 µl ice-cold TBST and once with 300 µl ice-cold WBT. Remove WBT only when the stopped translation reaction can be added to the wells (see Subheading 3.4).
4. Add the stopped *in vitro* translation, seal the plate, and incubate the binding reaction at 4°C for 1 h. Wash the wells with 300 µl ice-cold WBT containing 0.1% BSA for eight to ten times. Use two fast washes removing the buffer immediately, followed by incubations starting at 5 min and extending to 15 min in later rounds. In these longer incubations binders with fast off-rates will dissociate and subsequently be washed away.

5. For elution of the RNA, add 100 µl EB containing EDTA to release the mRNA from the captured protein–mRNA–ribosome complexes and freshly added *S. cerevisiae* RNA (final concentration 50 µg/ml) to block the surface of the tubes and perhaps to act as competing substrate for any residual RNases. Incubate at 4°C for 10 min and add to 400 µl lysis buffer of the High Pure RNA purification kit on ice. Repeat the elution step and collect the second elution in the same tube. After vortexing the RNA is stable and can be processed at room temperature until elution from the column (see Subheading 3.6).

3.5.3. Selection in Solution

1. Starting from the stopped and centrifuged *in vitro* translation reaction (see Subheading 3.4), divide the reaction into two aliquots of 250 µl and add 250 µl of STOP mix. Add 40 µl of streptavidin-coated magnetic beads that were washed two times with 500 µl TBS and blocked with 500 µl TBST/0.5% BSA for 1 h in a 2-ml tube as preselection step (see Note 14). Rotate at 4°C for 1 h.
2. Transfer the supernatant to a blocked 2-ml tube and add to 100–200 nM of biotinylated target (omit target in the control reaction) and incubate rotating at 4°C for 1 h (see Note 15).
3. Transfer the supernatant to a blocked tube containing 40 µl of streptavidin-coated magnetic beads and capture the ternary complexes rotating at 4°C for 30 min. Wash with 500 µl ice-cold WBT containing 0.1% BSA as indicated above (see Subheading 3.5.2). Separate captured complexes using a magnetic separator between each washing step.
4. Proceed with the elution and purification of RNA as described for the selection on plates (see Subheading 3.5.2).

3.5.4. Affinity Maturation by Competition with Non-labeled Target (Off-Rate Selection)

To increase the affinities of the library members, it is best to select for those having the lowest dissociation rate constant (off-rate) from the target (7, 13, 15, 28, 29). This off-rate selection can be applied for the improvement of known binders (after mutagenizing the gene for defined binders and thus creating a new library), but also during the initial selection from the original library. In this off-rate selection step an excess of nonbiotinylated target is added after the binding reaction has already been equilibrated for 1 h. Any fast dissociating binder will be immediately occupied by nonbiotinylated target and thereby prevented from being captured with biotinylated target on streptavidin or neutravidin. Conversely, any high-affinity binder with a slow off-rate will retain its biotinylated target and thus can be captured. The optimal duration of competitor incubation and the excess concentrations depend on the expected off-rates. As a general guideline, for the first off-rate selection a 2-h incubation with a 10 to 100-fold excess of competitor is recommended, and a 14-h incubation

with a 1,000 to 10,000-fold excess of competitor in later rounds may be appropriate. Then proceed with washing and elution of the bound ternary complexes as above.

3.6. Recovery of Eluted RNA

1. Apply the lysis buffer/eluate mixture on the column of the High Pure RNA isolation kit (see Note 16; *Optional:* as a positive control also purify 2 µl of the input RNA from the *in vitro* transcription diluted in 200 µl EB) and centrifuge at 8,000×*g* for 1 min. Discard the flow-through.
2. Add 100 µl diluted DNAse I solution (1.8 U/µl) directly onto the column filter and incubate at room temperature for 15 min (see Note 17). Add 500 µl wash buffer 1 and centrifuge at 8000×*g* for 1 min. Discard flow-through.
3. Wash with 500 µl wash buffer 2, centrifuge and discard flow-through.
4. Add 100 µl wash buffer 2 and centrifuge at 13,000×*g* for 2 min to remove any residual EtOH.
5. Elute with 50 µl elution buffer and incubate for 2 min before centrifugation at 8,000×*g* for 1 min into a fresh 1.5-ml RNAse-free tube.
6. Freeze the remaining sample of eluted RNA in liquid nitrogen and store at –80°C (see Notes 10 and 18).

3.7. Reverse Transcription of DARPin mRNA

1. Transfer two times 12.5 µl of eluted RNA to fresh 1.5-ml tubes (see Note 19).
2. Denature the eluted RNA at 70°C for 10 min and chill on ice.
3. Set up the following reverse transcription (RT)-mix (total of 7.75 µl) per RT reaction on ice:
   - 0.25 µl WTC4 primer (final concentration 1.25 µM)
   - 0.5 µl dNTPs (final concentration 125 µM of each nucleotide)
   - 0.5 µl RNasin (40 U/µl)
   - 0.5 µl AffinityScript™ Multiple Temperature Reverse Transcriptase (50 U/µl)
   - 2.0 µl 10× AffinityScript buffer
   - 2.0 µl DTT (final concentration 10 mM)
   - 2.0 µl H_2O
4. Distribute 7.75 µl RT-mix per RT reaction to the 12.25 µl samples of denatured RNA.
5. Incubate at 50°C for 1 h.
6. Use 2–5 µl as template for PCR using the inner primers WTC4 and EWT5s.
7. Freeze the rest of the cDNA in liquid nitrogen and store at –20°C.

3.8. Amplification of cDNA Coding for DARPins

The standard protocol for Vent DNA polymerase (NEB) is shown below. If another DNA polymerase or primers are used, the reaction conditions might have to be adapted.

1. Set up the following reaction mix per sample:
   - 2–5 µl cDNA
   - 5.0 µl 10× Thermopol buffer (NEB)
   - 2.0 µl dNTPs (final concentration 200 µM of each nucleotide)
   - 2.5 µl DMSO (final concentration 5%)
   - 0.5 µl WTC4 primer (final concentration 1 µM)
   - 0.5 µl EWT5s primer (final concentration 1 µM)
   - 0.5 µl Vent DNA Polymerase (2 U/µl)
   - Add to 50 µl with H_2O
2. Perform a hot start PCR to reduce unspecific amplification. Use the following cycling parameters: 3 min at 95°C, 25 cycles: 30 s at 95°C, 30 s at 55°C, 45 s at 72°C, and final extension 5 min at 95°C (see Note 20).
3. Verify the product on an agarose gel (see Note 21).

3.9. Incorporation of Promoter Elements, RBS, tolA Spacer and RNA-Stabilizing Stem Loops

1. Purify PCR products either by excision of the according bands from the agarose gel and subsequent purification or direct purification over commercially available columns, for example, of the NucleoSpin extract II kit. Elute in a small volume of 20 µl.
2. Digest≥150 ng of the PCR product with the corresponding restriction enzymes, e. g., *Nco*I and *Hin*dIII for DARPin selections, in a final volume of 30 µl at 37°C for 2 h (see Note 22).
3. Purify digested PCR product using the NucleoSpin® Extract II DNA purification kit. Elute in 15 µl elution buffer supplied with the kit.
4. Ligate the PCR fragments into the ligation-ready pRDV plasmid using 100 ng of digested pRDV and the digested PCR product with a molar ratio of vector to insert of 1:5–7 in a final volume of 10 µl. Add 1 U of T4 DNA ligase and 1 µl ligase buffer. Incubate 30–60 min at room temperature. Use this ligation as PCR template with the T7B and tolAk primers (see Subheading 3.2) or perform an error-prone PCR to increase diversity (see Subheading 3.10).

3.10. Error-Prone PCR

1. Set up reactions introducing different mutational rates using various concentrations in the range of 1–20 µM of the nucleotide analogs dPTP and 8-oxo-dGTP (see Note 23): 10 ng pRDV_DARPin template, 250 µM dNTPs each, 1 µM T7B primer and tolAk primer, 1× polymerase buffer, 1.5 mM $MgCl_2$ and 0.2 µl Platinum® Taq DNA Polymerase in a 50 µl reaction.

2. Apply the following cycling parameters (must be adapted according to primers and template): 3 min at 95°C, 25 cycles: 30 s at 95°C, 30 s at 50°C, 1 min at 72°C, and final extension 5 min at 95°C.
3. Verify the product on an agarose gel.
4. Mix PCR products in equimolar amounts to serve as template for the *in vitro* transcription (see Subheading 3.2, step 4).

3.11. Initial Analysis of Selected Individual Library Members in a 96-Well Format (22)

1. After RT-PCR (see Subheadings 3.7 and 3.8) prepare the DARPin pool after enrichment has been observed for subcloning into a prokaryotic expression plasmid using the endonucleases *Bam*HI and *Hin*dIII. Enrichment is indicated by a much stronger PCR band recovered from a well with immobilized target than from a control well without immobilized target.
2. Ligate the PCR fragment into pDST67 (20, 25) as fusion with the sequence coding for a N-terminal $MRGS(H)_6$ tag for purification.
3. After transformation into *E. coli* XL1-Blue pick single clones and inoculate in deep 96-well plates in 1 ml 2×TY/1% glucose/amp (100 µg/ml) grow overnight at 37°C while shaking at 540 rpm on an orbital shaker.
4. Transfer 100 µl of each culture to 900 µl fresh media and grow at 540 rpm for 1 h at 37°C.
5. Induce with 0.5 mM IPTG (add 100 µl media containing 5.5 mM IPTG) and grow an additional 3–5 h at 37°C.
6. Harvest cells by centrifugation at 400×*g* for 10 min, and discard supernatant.
7. Resuspend pellet in 50 µl B-PER II detergent and lyse cells for 15–30 min on an orbital shaker at 500 rpm.
8. Add 1 ml TBST/0.1% BSA and centrifuge to remove debris.
9. Use 10 µl (using a predilution of 1:100 in TBST/0.1% BSA can give you a better indication of the affinity of the binders when they are in the µM to low nM range) for ELISA. Binders with even higher affinity will still fully saturate the immobilized target and can only be distinguished in ELISA by inhibition with low concentrations of soluble target.
10. For ELISA, coat wells with 100 µl of 66 nM neutravidin in TBS for 1 h at room temperature or overnight at 4°C. Wash two times with 300 µl TBS. Dry plate on paper towels after each step.
11. Block with 300 µl TBST/0.5% BSA for 1 h at room temperature.
12. Invert plate and shake out liquid and immediately add 100 µl of the biotinylated target (10–100 nM) in TBST/0.1% BSA. Incubate 1 h at 4°C or room temperature on an orbital shaker (see Note 24). Wash three times with 300 µl TBST.

13. Add 100 µl DARPin extract from step 9. Incubate 1 h at 4°C or room temperature on an orbital shaker. Wash three times with 300 µl TBST.
14. Add 100 µl mouse-anti-RGS(His)$_4$ antibody in a 1:500 dilution. Incubate 1 h at 4°C or room temperature on an orbital shaker. Wash three times with 300 µl TBST.
15. Add 100 µl goat-anti-mouse antibody coupled to alkaline phosphatase in a 1:20,000 dilution. Incubate 1 h at 4°C or room temperature on an orbital shaker. Wash three times with 300 µl TBST.
16. Add 100 µlpNPP substrate solution. Incubate until color development at 4° to 37°C depending on the stability of the target and the library scaffold and determine $OD_{405\ nm}$.

3.12. Preparation of S30 Extract (30–32)

1. Grow a 100-ml culture of *E. coli* MRE600 in incomplete rich medium overnight at 37°C.
2. Transfer 10 ml of the overnight culture in 1 l of fresh media in a 5-l baffled shaker flask and grow until $OD_{600\ nm}$ of 1.0–1.2 at 37°C while shaking at 230 rpm. This procedure can be scaled up to your needs, and 1 l culture usually yields 10–15 ml of S30 extract. The S30 extract is stable for years when stored at –80°C.
3. Chill cultures for 10 min on an ice water bath with gentle shaking.
4. Centrifuge cells at 3,500 × *g* at 4°C for 15 min and discard supernatant.
5. Wash the pellet three times with 50 ml of ice-cold S30 buffer per 1 l culture. It is best to resuspend cells with plating beads or on a magnetic stirrer using a sterile stir bar.
6. Freeze the cell pellet in liquid nitrogen and store for a maximum of 2 days at –80°C or continue immediately.
7. Resuspend the cell pellet (use 50 ml ice-cold S30 buffer per 1 l of culture), centrifuge at 4,000 × *g*. Discard supernatant and resuspend pellet in 4 ml S30 buffer per g wet cells (typically 1 l of culture yields 1.5–2.0 g cell pellet).
8. Lyse the cells by one single passage through a French press applying 1,000 psi or an EmulsiFlex at ~17,000 psi.
9. Centrifuge cells at 20,000 × *g* (SS-34) at 4°C for 30 min. Transfer supernatant to clean centrifuge bottle(s) and repeat this step.
10. Add 1 ml of preincubation mix to each 6.5 ml of cleared supernatant (usually 1 l culture will yield 8–10 ml of S30 extract) and slowly shake at 25°C for 1 h. In this time all

endogenous mRNA will be translated and cellular nucleases will degrade mRNA and DNA (33).

11. Dialyze the S30 extract in a tubing with a MW cutoff of 6,000-8,000 Da (Spectrum Laboratories SpectraPor No. 132650) against a 50-fold volume of S30 buffer at 4°C three times for 4 h.
12. Centrifuge S30 extract at 4,000 × *g* at 4°C for 10 min. If the library members and target are devoid of disulfide bonds, 1 mM DTT can be added to the extract. Aliquot at 4°C in suitable volumes (e.g., 55 µl is sufficient for one *in vitro* translation reaction, 110 µl for two) since it should not be refrozen to guarantee best activity. Flash freeze in liquid nitrogen and store at –80°C.

4. Notes

1. Use RNAse-free water, chemicals, and consumables. Most commercially available water is RNase-free or can be generated using a membrane microfiltration system, e.g., MilliQ from QIAGEN, to produce ultrapure water. Alternatively, you can use 0.1% DEPC (diethylpyrocarbonate), which reacts with histidine residues but also other nucleophilic groups and therefore inactivates RNases, but for the same reason it cannot be used, for example, with Tris-buffers. Chemicals should be kept separate from the common chemical shelf and handled only with gloves and a flamed spatula to avoid RNase contamination. Purchase only RNase-free plastic consumables. If necessary you can bake glass bottles and pipettes at 180°C for 6 h.
2. The buffer composition may be adjusted to the requirements of the library and target, but it is important that the wash buffer contains 50 mM Mg^{2+} to stabilize the ribosome. It is recommended to test buffer conditions with a known binder to ensure stability of the nascent chain complex.
3. Different reverse transcriptases [AffinityScript™ Multiple Temperature Reverse Transcriptase, SuperScript™ II (Invitrogen, No. 18064-022), ThermoScript™ (Invitrogen, No. 12236-014), and QuantiTect (QIAGEN, No. 205310)] were tested for efficiency on DARPin sequences. With exception of QuantiTect the yield obtained was comparably high with all other reverse transcriptases.
4. Previously Phusion™ High-Fidelity DNA Polymerase (New England Biolabs, No. F-530S) has also been used (26, 30). Different DNA polymerases were tested [$Vent_R$® DNA Polymerase, Herculase® II Fusion DNA Polymerase (Stratagene,

No. 600677), Expand High Fidelity PCR System (Roche Diagnostics, No. 11732641001]. The DNA polymerase mix from the Expand High Fidelity PCR System gave the lowest yield of PCR product, while Herculase II gave the highest amount of side products. Therefore, we now routinely use $Vent_R^®$ DNA Polymerase for amplification of DARPin sequences. Since the yield was highest with the Herculase II it might be a good alternative to increase the yield of PCR product or for amplification of other library scaffolds.

5. Use the home-made RNA polymerase buffer (see Subheading 2.4) as indicated for maximum yield of RNA. Commercial buffers have not worked very well at this step when the PCR product is directly used without further purification.
6. In round one ensure that the number of molecules actually exceeds the library size. Conversely, in a newly constructed library, the diversity cannot be higher than the number of molecules used in this step. In later rounds, an enrichment is obtained, and it is generally sufficient to use ~50 ng of pRDV_DARPin template.
7. The PCR products can be used without additional purification.
8. Optionally, the transcribed RNA can be analyzed on a denaturing formaldehyde agarose gel following standard procedures (34). The mRNA product should give a sharp band. A smear or no product indicates RNase contamination, which needs to be eliminated and the step repeated. If the band is sharp but the yield is lower than expected: Obtain more starting DNA template by not purifying the PCR product that is used as template, as the quality is usually sufficient even without purification, and do use the home-made RNA polymerase buffer (Subheading 2.4) for better transcription yield. If the products are not of the expected size, optimize the PCR conditions depending on your template and primers.
9. In some cases the template DNA might bind unspecifically to the target, for example, if the target is highly positively charged, and then it is recommended to remove this contamination by DNAse I treatment before the actual selection. Always freeze small aliquots of DNAse I and store at –20°C. Do not refreeze or vortex solutions containing DNAse I, because the enzyme is very sensitive to denaturation.
10. Always freeze RNA immediately after use and only thaw when needed to avoid degradation.
11. For the selection, some general considerations need to be pointed out. Always use the same target preparation through all of the selection and screening rounds, and ensure its quality and account for its stability over the duration of the experiment. If the target denatures, epitopes present in the native

protein will vanish, and such binders will be lost. Account for high diversity especially in the first round by using sufficient starting library. Start selections with a higher number of DNA template molecules than the diversity of the library. To extract all putative binders in the library use a larger surface area to immobilize the target in the first round. The first round should, in general, not be highly selective; it is more important to capture the full diversity of binders, as a binder lost at this stage can never be recovered. In general, it is recommended to perform the selection in duplicates to monitor the selection quality. It is recommended to switch between neutravidin (a chemically modified derivative of avidin) and streptavidin, or even switch between selections on immobilized target and target in solution during the selection process, to focus selection on binding to the target, rather than on streptavidin/neutravidin or any other surface features. If high-affinity binders in the pM range are needed, include the introduction of additional random mutations using error-prone PCR and increase stringency by applying off-rate selections. Perform one cycle of nonstringent selection including an error-prone PCR followed by a round of off-rate selection without error-prone PCR (see Subheading 3.5.4). The rationale is that error-prone PCR will generate many nonfunctional molecules. First, *all* functional molecules should be recovered by a nonstringent selection, then from this pool of functional (randomized) molecules, the best ones should be recovered. Use these to perform a stringent round using off-rate selection. Start at 10–100-fold excess competitor; increase to 100–10,000-fold in later rounds, if feasible (35). These two rounds should be followed again by a nonstringent round without any additional selection pressure; simply to amplify the rare molecules. Perform this cycle of error-prone PCR, off-rate selection, and nonstringent round two to three times before analyzing single clones (see Subheading 3.10).

12. Using the AviTag has the advantage that all biotinylated proteins are labeled uniformly and remote from epitopes, which might interfere with later use and are labeled only once, leading to a more homogenous target preparation. Avoid the presence of a Met-Arg-Gly-Ser-$(His)_6$ tag ("RGS-His-tag") on the biotinylated target, rather use a $(His)_6$ tag for purification, since the detection of DARPins bound to the target is performed using an anti-RGS$(His)_4$ antibody (see Subheading 3.10). Make sure your target sample is devoid of free biotin. Biotin removal requires an extensive dialysis, for example, four times against an 100-fold volume buffer for 4 h each. Nonbiotinylated target can be removed using a monomeric avidin column following the manufacturer's instructions (Pierce, No. 53146).

13. Use one well as nontarget control and two wells with immobilized target in later rounds as mutual controls for enrichment. When starting from the libraries in round 1 it is recommended to use a larger surface, for example, four wells with immobilized target.
14. To remove unspecifically binding ribosomal complexes it is recommended to use a preselection on BSA-blocked wells coated only with neutravidin or streptavidin, but omitting the target protein, except for round 1, where this "prepanning" should not be done. For prepanning, the preparation of additional wells and incubation of the ternary complexes from the *in vitro* transcription for 30–60 min are necessary before transferring the solution to the target-coated or control wells.
15. The amount of target can be reduced to 100 pM, for example, when performing an off-rate selection, and thus a high amount of competitor can be added. At even lower target concentrations the unspecific binding might dominate over target binding, however, and thus specificity of binding must be carefully controlled.
16. RNA isolation can also be performed with the RNeasy mini kit (QIAGEN, No. 74104) with comparable yield of resulting PCR product.
17. This step is highly recommended to avoid amplification of nonselected template DNA that has been carried over through all steps of the selection procedure. See also Note 9 for handling of DNAse I.
18. The RNA should be stable for years at –80°C, but we recommend to immediately proceed with cDNA synthesis and PCR amplification for best recovery of sequences of putative binders.
19. Use one sample without addition of reverse transcriptase as control. The result of the following PCR will be a measure for the quality of the selection regarding DNA carry-over from the input DNA and putative overcycling (see Notes 17 and 20).
20. Depending on the round of selection more or fewer cycles could be advantageous. In the first round, 32–40 cycles are recommended to obtain sufficient product. After more rounds of selection specific binders are being enriched, therefore, the output of eluted RNA molecules increases. By lowering the cycle numbers in round 2 to between 28 and 35 and in all following rounds to 25, unspecific amplification can be reduced to a minimum. In addition, note that when the selection pressure increases, for example, after off-rate selection, the yield of PCR product might decrease. In this case use more cycles.
21. If the quality and amount (<10 ng/μl) of the PCR product was not satisfactory repeat the PCR. Never reamplify the PCR product, because this might lead to unspecific amplification of unwanted by-products.

22. In parallel, digest the ribosome display vector (pRDV) with the same restriction enzymes, for example, *NcoI* and *Hin*dIII for DARPin selections. Purify the plasmid backbone using extraction from a preparative agarose gel. It is recommended to use a larger preparation to last for several selection rounds and/or multiple target selections: Test the quality of the digested plasmid by ligation and transformation and/or PCR on the ligation mix to evaluate the level of religation and therefore quality of the ligation-ready plasmid.
23. Add different concentrations of the nucleotide analogs, for example, 0, 3, and 10 μM. Up to 20 μM can be used, but the amount of product is greatly reduced at this concentration. The mutational load per kb under the conditions described is 1.5 mutations with 1 μM nucleotide analogs and 3.2 mutations with 3 μM nucleotide analogs using Platinum Taq polymerase. These numbers refer to fresh nucleotides and can vary if the nucleotides are no longer incorporated well, for example, by hydrolysis of the triphosphate. The use of a low to medium mutational load per selection, but repeating over several rounds, might be beneficial over a high mutational load which might result in a high number of misfolded library members in the pool.
24. The incubation temperature depends on the stability of the target and library scaffold.
25. Some applications, e.g., for therapy, require high stability of the therapeutic agent (36). Selection for high stability can also be achieved with ribosome display. This is best achieved by first making the whole population unable to fold, by introducing a reversible destabilization, and then selecting for compensating mutations, and finally removing the destabilization again. For example, most antibody domains require disulfides for stability, which form only under oxidizing conditions. A destabilization and increase in aggregation of the antibody fold is usually observed when the disulfides are removed (6, 37). Using a reducing environment during the selection, scFv antibody fragments could be evolved that were able to fold under reducing conditions correlating with conditions in the cytosol, and they showed higher stability than the starting molecule in the absence of the disulfide bonds (7), but also after the disulfide bonds were allowed to form again. Antibody fragments with these improved biophysical property can be used in biomedical applications with disulfides formed, but they also make an intracellular application (as “intrabodies”) (38) more feasible. In addition, rational design of the antibody framework (39) could contribute to the development of stability-improved, antibody-based therapeutics.

Acknowledgments

We thank many former and current members of the Plückthun laboratory for establishing and continuously optimizing the ribosome display protocol. In this respect we thank the following people not only, but especially for the following efforts: Petra Parizek, Birgit Lindner, and Gabriela Nagy-Davidescu for revision of the pRDV plasmid sequence, Oliver Scholz for the testing of the EWT5s primer, amplification conditions, and RNA yield, Fabio Parmeggiani for testing of the illustra G50 columns, and Gautham Varadamsetty and Hilmar Ebersbach for their help in testing enzymes on different protein scaffolds. Work on ribosome display was supported by the Swiss National Science Foundation.

References

1. Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**,1315–17.
2. Hanes, J. and Plückthun, A. (1997) *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. USA* **94**, 4937–42.
3. Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R., and Plückthun, A. (1998) Ribosome display efficiently selects and evolves high-affinity antibodies *in vitro* from immune libraries. *Proc. Natl. Acad. Sci. USA* **95**, 14130–5.
4. Roberts, R. W. and Szostak, J. W. (1997) RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci. USA* **94**, 12297–302.
5. Liu, R., Barrick, J. E., Szostak, J. W., and Roberts, R. W. (2000) Optimized synthesis of RNA-protein fusions for *in vitro* protein selection. *Methods Enzymol.* **318**, 268–93.
6. Wörn, A. and Plückthun A. (2001) Stability engineering of antibody single-chain Fv fragments. *J. Mol. Biol.* **305**, 989–1010.
7. Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J., and Plückthun, A. (2001) Tailoring *in vitro* evolution for protein affinity or stability. *Proc. Natl. Acad. Sci. USA* **98**, 75–80.
8. Stemmer, W. P. (1994) Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389–91.
9. Zaccolo, M., Williams, D. M., Brown, D. M., and Gherardi, E. (1996) An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* **255**, 589–603.
10. Mattheakis, L. C., Bhatt, R. R., and Dower, W. J. (1994) An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. USA* **91**, 9022–6.
11. Hanes, J., Jermutus, L., and Plückthun, A. (2000) Selecting and evolving functional proteins *in vitro* by ribosome display. *Methods Enzymol.* **328**, 404–30.
12. Hanes, J., Schaffitzel, C., Knappik, A., and Plückthun, A. (2000) Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat. Biotechnol.* **18**, 1287–92.
13. Luginbühl, B., Kanyo, Z., Jones, R. M., Fletterick, R. J., Prusiner, S. B., Cohen, F. E., Williamson, R. A., Burton, D. R., and Plückthun, A. (2006) Directed evolution of an anti-prion protein scFv fragment to an affinity of 1 pM and its structural interpretation. *J. Mol. Biol.* **363**, 75–97.
14. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P., and Plückthun, A. (2003) Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489–503.
15. Zahnd, C., Wyler, E., Schwenk, J. M., Steiner, D., Lawrence, M. C., McKern, N. M., Pecorari, F., Ward, C. W., Joos, T. O., and Plückthun, A. (2007) A designed ankyrin repeat protein evolved to picomolar affinity to Her2. *J. Mol. Biol.* **369**, 1015–28.

16. Binz, H. K., Amstutz, P., Kohl, A., Stumpp, M. T., Briand, C., Forrer, P., Grütter, M. G., and Plückthun, A. (2004) High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**, 575–82.
17. Amstutz, P., Binz, H. K., Parizek, P., Stumpp, M. T., Kohl, A., Grütter, M. G., Forrer, P., and Plückthun, A. (2005) Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins. *J. Biol. Chem.* **280**, 24715–22.
18. Zahnd, C., Pécorari, F., Straumann, N., Wyler, E., and Plückthun, A. (2006) Selection and characterization of Her2 binding-designed ankyrin repeat proteins. *J. Biol. Chem.* **281**, 35167–75.
19. Schweizer, A., Roschitzki-Voser, H., Amstutz, P., Briand, C., Gulotti-Georgieva, M., Prenosil, E., Binz, H. K., Capitani, G., Baici, A., Plückthun, A., and Grütter, M. G. (2007) Inhibition of caspase-2 by a designed ankyrin repeat protein: specificity, structure, and inhibition mechanism. *Structure* **15**, 625–36.
20. Huber, T., Steiner, D., Röthlisberger, D., and Plückthun, A. (2007) *In vitro* selection and characterization of DARPins and Fab fragments for the co-crystallization of membrane proteins: The Na^+-citrate symporter CitS as an example. *J. Struct. Biol.* **159**, 206–21.
21. Schatz, P. J. (1993) Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in *Escherichia coli*. *Biotechnology* **11**, 101138–43.
22. Wade, H. E. and Robinson, H. K. (1966) Magnesium ion-independent ribonucleic acid depolymerases in bacteria. *Biochem. J.* **101**, 467–79.
23. Laminet, A. A. and Plückthun, A. (1989) The precursor of beta-lactamase: purification, properties and folding kinetics. *EMBO J.* **8**, 1469–77.
24. O'Callaghan, C. H., Morris, A., Kirby, S. M., and Shingler, A. H. (1972) Novel method for detection of beta-lactamases by using a chromogenic cephalosporin substrate. *Antimicrob. Agents Chemother.* **1**, 283–8.
25. Steiner, D., Forrer, P., and Plückthun, A. (2008) Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display. *J. Mol. Biol.* **382**, 1211–27.
26. Zahnd, C., Amstutz, P., and Plückthun, A. (2007) Ribosome display: selecting and evolving proteins *in vitro* that specifically bind to a target. *Nat. Methods* **4**, 269–79.
27. Hajnsdorf, E., Braun, F., Haugel-Nielsen, J., Le Derout, J., and Régnier, P. (1996) Multiple degradation pathways of the rpsO mRNA of *Escherichia coli*. RNase E interacts with the 5′ and 3′ extremities of the primary transcript. *Biochimie* **78**, 416–24.
28. Hawkins, R. E., Russell, S. J., and Winter, G. (1992) Selection of phage antibodies by binding affinity. Mimicking affinity maturation. *J. Mol. Biol.* **226**, 889–96.
29. Yang, W. P., Green, K., Pinz-Sweeney, S., Briones, A. T., Burton, D. R., and Barbas, C. F. 3rd. (1995) CDR walking mutagenesis for the affinity maturation of a potent human anti-HIV-1 antibody into the picomolar range. *J. Mol. Biol.* **254**, 392–403.
30. Amstutz, P., Binz, H. K., Zahnd, C., and Plückthun, A. (2006) Ribosome display: *In vitro* selection of protein-protein interactions. In "*Cell Biology: A Laboratory Handbook*" (Celis, J., ed.), 497–509. Elsevier, Amsterdam.
31. Chen, H. Z. and Zubay, G. (1983) Prokaryotic coupled transcription-translation. *Methods Enzymol.* **101**, 674–690.
32. Pratt, J. M. (1984) Coupled transcription-translation in prokaryotic cell-free systems. In *"Current Protocols"* (Hemes B. D. and Higgins S. J., eds.), 179–209. IRL Press, Oxford.
33. Kushner, S. R. (2002) mRNA decay in *Escherichia coli* comes of age. *J. Bacteriol.* **184**, 4658–65; discussion 4657.
34. Sambrook, J. and Russel, D. W. (2001) Molecular cloning: A laboratory handbook, 3rd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
35. Zahnd, C., Sarkar, C. A., and Plückthun, A. (2010) Computational analysis of off-rate selection experiments to optimize affinity maturation by directed evolution. *Protein Eng. Des. Sel.* **23**, 175–84.
36. Willuda, J., Honegger, A., Waibel, R., Schubiger, P. A., Stahel, R., Zangemeister-Wittke, U., and Plückthun, A. (1999) High thermal stability is essential for tumor targeting of antibody fragments: Engineering of a humanized anti-epithelial glycoprotein-2 (epithelial cell adhesion molecule) single-chain Fv fragment. *Cancer Res.* **59**, 5758–67.
37. Proba, K., Wörn, A., Honegger, A., and Plückthun, A. (1998) Antibody scFv fragments without disulfide bonds made by molecular evolution. *J. Mol. Biol.* **275**, 245–53.
38. Chames, P. and Baty D. (2000) Antibody engineering and its applications in tumor targeting and intracellular immunization. *FEMS Microbiol. Lett.* **189**, 1–8.
39. Honegger, A., Malebranche, A. D., Röthlisberger, D., and Plückthun, A. (2009) The influence of the framework core residues on the biophysical properties of immunoglobulin heavy chain variable domains. *Protein Eng. Des. Sel.* **22**, 121–34.

Chapter 22

GLOBE: Analysis of DNA–Protein Interaction Analysis

Takaaki Kojima and Hideo Nakano

Abstract

Emulsion PCR, a hyper, multi, parallel PCR in water-phase droplets in water-in-oil (w/o) emulsion, can be used to make a *g*enetic *l*ibrary *o*n *be*ads (GLOBE). In GLOBE, we have developed a novel high-throughput screening system for the analysis of the recognition sequences of DNA-binding proteins, which can be prepared by using either an in vivo or an in vitro protein synthesis system. The system can contribute to the low-cost comprehensive analysis of transcription factor-binding regions.

Key words: Emulsion PCR, Cell-free protein synthesis system, Transcription factor, Flow cytometry, GLOBE

1. Introduction

Gene regulation is a complex but central process in all organisms. Determining the binding sites of regulatory proteins in genomes is important for understanding transcriptional regulatory networks (1). The binding of a transcription factor to its genomic targets can be assayed by chromatin immunoprecipitation (ChIP) (2). A new assay, known as "ChIP-chip," combines the ChIP method with microarray (chip) hybridization (3). The ChIP-chip method is extremely powerful in identifying transcription factor targets; however, it can be applied only when a specific antibody to the transcription factor and promoter chip of target organisms are available.

An in vitro ligand evolution system for systematic evolution of ligands by exponential enrichment (SELEX) has been applied for the analysis of target sequence of transcription factors; SELEX-SAGE (SELEX and serial analysis of gene expression) was successfully used for the sequencing of more than 10,000 potential DNA sequences for the CTF/NFI transcription factor (4).

Daniel.J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_22, © Springer Science+Business Media, LLC 2011

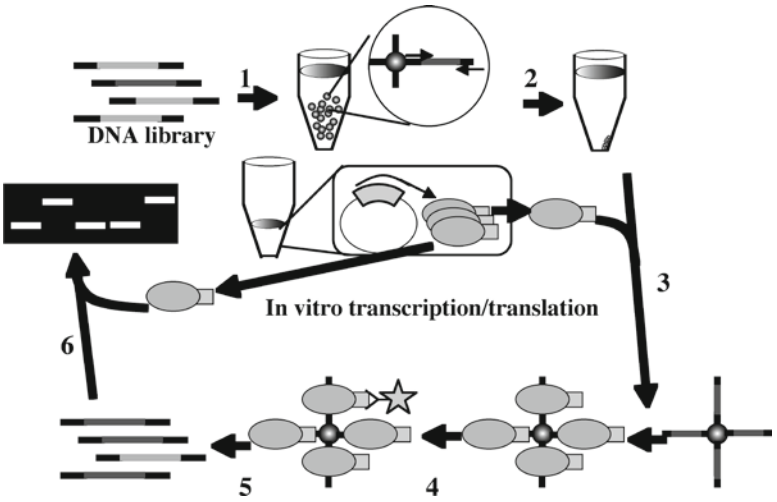


Fig. 1. Schematic drawing of screening assay of DNA–protein interaction with *g*enetic ibrary n ads (GLOBE). (1) GLOBE construction with solid-phase single-molecule PCR in water-in-oil (w/o) emulsions. (2) Emulsions are broken and beads are recovered. (3) In vitro-expressed DNA-binding protein with epitope tag is added to the beads. (4) A fluorescence-labeled anti-tag antibody is added to the complex. (5) Flow cytometry for selecting positive clones and bead PCR with selected beads. (6) Gel-mobility shift assay linked with in vitro transcription/translation (reproduced from ref. 6).

However, the SELEX method needs a highly purified DNA-binding protein.

Recently, we have devised a novel method to make a *g*enetic *l*ibrary *o*n *be*ads (GLOBE) by using solid-phase single-molecule PCR in water-in-oil (w/o) emulsions (emulsion PCR). Using GLOBE, we developed a high-throughput screening system for DNA–protein interactions (Fig. 1) and showed that it is a powerful tool for screening the binding regions of transcription factors (5, 6). In this chapter, we describe the protocols of the emulsion PCR and the high-throughput analysis of DNA–protein interactions by flow cytometry.

2. Materials

2.1. Oligonucleotides

1. pBlue-Reverse-NH_2: NH_2-TTTTTTTTGG AAACAGCTAT GACCAT.
2. M13M1cc: CCAGTCACGA CGTTGTA.
3. pBRN: TTTTTTTTGG AAACAGCTAT GACCAT.

4. RF-Xba: GCTCTAGATA AGAAGGAGAT ATACATATGG CTGAGCC-CGA GAACACGA.
5. RR-Bam: CGGGATCCTG TTAGCAGCCG GATCTCA-GTG G.
6. F1 primer: ATCTCGATCC CGCGAAATTA ATACG.
7. R1 primer: TCCGGATATA GTTCCTCCTT TCAG.
8. pBlue-Reverse-bio: biotin-GGAAACAGCT ATGACCAT.
9. pBRN-Cy3: Cy3-GGAAACAGCT ATGACCATGA G.
10. Pp-1-F-N9:GGAAACAGCTATGACCATGAGCTCGTAAAA TTTTTCTGCN NNNNNNNNAG AATACAACGT CGTGACTGG (number of N is a variable depending on each transcription factor).

2.2. Genetic Library on Beads

1. Oligonucleotides (Pp-1-F-N9 and M13M1cc).
2. Klenow fragment reaction mixture: 10 mM Tris–HCl (pH 7.5), 7 mM $MgCl_2$, 0.1 mM DTT, and 0.2 mM of each dNTP.
3. 2 U/μL Klenow fragment (Takara).
4. Phenol–chloroform: TE-saturated phenol/chloroform/isoamyl alcohol (25:24:1).
5. Ethanol and 3 M sodium acetate (adjusted to pH 5.2 using acetic acid).
6. Recovery buffer: 500 mM ammonium acetate and 1 mM EDTA (adjusted to pH 8.0 using NaOH).
7. TE buffer: 10 mM Tris–HCl (pH 8.0) and 1 mM EDTA (adjusted to pH 8.0 using NaOH).
8. Magnetic beads (Dynabeads M-270 carboxylic acid, 2.8 μm ± 0.2 in diameter, Invitrogen).
9. 2-(*N*-morpholino) ethane sulfonic acid (MES) buffer (0.4 M, adjusted to pH 5.0 using NaOH).
10. 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC).
11. PCR mixture (water phase): 1× *ExTaq* buffer (Takara), 0.2 mM of each dNTP, 0.25 μM M13M1cc, random DNA library (0.33 molecules per compartment on average), 8×10^6 pBlue-Reverse-NH_2-coupled beads, 2.5 nM pBRN, and 0.25 U/μL *ExTaq* DNA polymerase (Takara).
12. Oil phase [Mineral oil (Sigma) containing 1% Sun Soft No. 818SK (Taiyo Kagaku)] (see Note 1).
13. B&W buffer: 1 M NaCl, 5 mM Tris–HCl (pH 8.0), and 0.5 mM EDTA (adjusted to pH 8.0 using NaOH).

2.3. Cell-Free Protein Synthesis

1. Expression vector of His-tagged transcription factor: pRSET-phaR-His (6) (see Note 2).

2. PCR mixture: 1× *ExTaq* buffer (Takara), 0.2 mM of each dNTP, 0.5 μM of each primer (F1 and R1 primers), and 0.025 U/μL *ExTaq* DNA polymerase (Takara).
3. Phenol–chloroform: TE-saturated phenol/chloroform/isoamyl alcohol (25:24:1).
4. Ethanol and 3 M sodium acetate (adjusted to pH 5.2 using acetic acid).
5. In vitro-coupled transcription/translation reaction mixture: template DNA for in vitro expression (including T7 promoter, ribosome-binding site, His-tagged transcription factor gene, and T7 terminator) (~0.3 pmol), 56.4 mM Tris-acetate (pH 7.4), 1.22 mM ATP, 0.85 mM each of GTP, CTP, UTP, 50 mM creatine phosphate, 0.15 mg/mL creatine kinase, 0.5 mM each of 20 amino acid, 4% polyethylene glycol 6000, 34.6 μg/mL folinic acid, 0.17 mg/mL *Escherichia coli* tRNA, 35.9 mM ammonium acetate, 10 mM $Mg(OAc)_2$, 100 mM KOAc, 10 μg/mL rifampicin, 7.6 μg/mL T7 RNA polymerase, 5.4 mM dithiothreitol (DTT), and 28.3% *E. coli* S30 extract from a double protease-deficient mutant strain (Δ*clpP-ompT*) (7) (see Note 3).

2.4. Screening of GLOBE Using Flow Cytometry

1. Fluorescent-activated cell sorter [EPICS ELITE ESP (Beckman Coulter)].
2. PBS buffer: 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, and 1.4 mM KH_2PO_4 (pH 7.3).
3. Anti-His(C-term)-FITC (Invitrogen).
4. Sheath buffer [IsoFlow (Beckman Coulter)].

2.5. PCR Amplification from Beads–DNA

1. DNA dilution buffer (TE): 10 mM Tris–HCl (pH 8.0) and 1 mM EDTA (adjusted to pH 8.0 using NaOH).
2. PCR mixture: 1× *ExTaq* buffer (Takara), 0.2 mM of each dNTP, 0.25 μM of each primer (M13M1cc and pBRN), 0.025 U/μL *ExTaq* DNA polymerase (Takara).
3. Electrophoresis buffer (1× Tris–borate–EDTA, TBE): 89 mM tris-(hydroxymethyl)aminomethane, 89 mM boric acid, and 2 mM EDTA.
4. Electrophoresis gel: 1× TBE containing 2.5% agarose.
5. Phenol–chloroform: TE-saturated phenol/chloroform/isoamyl alcohol (25:24:1).
6. Ethanol and 3 M sodium acetate (adjusted to pH 5.2 using acetic acid).

2.6. Sequencing Analysis

1. ExoSAP reaction mixture: 1× *ExTaq* buffer (Takara), and 0.5 μL of ExoSAP-IT (GE Healthcare) in a total volume of 10 μL.

2. Sequencing primer: pBlue-Reverse-bio (see Note 4).
3. DNA-sequencing reagents: BigDye Terminator Cycle Sequencing kit (Applied Biosystems).

2.7. Gel-Mobility Shift Assay

1. Phenol–chloroform: TE-saturated phenol/chloroform/isoamyl alcohol (25:24:1).
2. Ethanol and 3 M sodium acetate (adjusted to pH 5.2 using acetic acid).
3. PCR mixture: 1× *Pyrobest* buffer II (Takara), 0.2 mM of each dNTP, 0.25 μM of each primer (M13M1cc and pBRN-Cy3), and 0.05 U/μL *Pyrobest* DNA polymerase (Takara).
4. TE buffer: 10 mM Tris–HCl (pH 8.0) and 1 mM EDTA (adjusted to pH 8.0 using NaOH).
5. Polyacrylamide gel: 12.5% polyacrylamide and 0.5× Tris–borate–EDTA (TBE).
6. 6× Loading buffer: 25% Glycerol, 5 mM Tris–HCl (pH 8.0), 0.5 mM EDTA (adjusted to pH 8.0 using NaOH), and 0.01% bromophenolblue.
7. Electrophoresis buffer (0.5× TBE).
8. Typhoon 9400 (GE Healthcare).

3. Methods

3.1. Construction of GLOBE

1. Magnetic bead solution (100 μL, 2×10^9 beads/mL) (Dynabeads M-270 carboxylic acid, Invitrogen) is washed twice with 0.01 N NaOH by mixing using a rotator at room temperature. The beads are washed three times with sterile water (S.W.) in the same manner and the supernatant is discarded.
2. Then, 20 μL of 5 -amino-modified oligonucleotide, pBlue-Reverse-NH_2 (100 μM), and 50 μL of MES buffer are added to the beads and the suspension is mixed using a rotator at room temperature.
3. After 30 min, 30 μL of MES buffer (including 3 mg EDC) is added to the bead solution (see Note 5). Then, the solution is mixed using a rotator for another 5 h at room temperature.
4. The beads are washed four times using 200-μL TE buffer; thereafter, the beads are suspended in 50-μL TE buffer and stored at 4°C.
5. 200 pmol of Pp-1-F-N9 and M13M1cc are added to the Klenow fragment reaction mixture, ensuring a total volume of 98 μL. These oligonucleotides are then denatured at 95°C for 3 min and annealed at room temperature (see Note 6).

6. A random DNA library is then synthesized by adding 4 U of Klenow fragment (Takara) to the product and incubating it at 37°C for 1 h.
7. The products are extracted using phenol–chloroform, precipitated with ethanol, and dissolved in sterile water (S.W.)
8. Then, the DNA library is separated by polyacrylamide gel electrophoresis, and then recovered from the gel. Once, the gel containing the DNA library is frozen at –80°C for 30 min.
9. After thawing, recovery buffer is added to the gel and mixed with a vortex mixer thoroughly for 2 h.
10. After the spin down, the supernatant is recovered. Purified DNA library extracted using phenol–chloroform, precipitated with ethanol, and dissolved in TE buffer and stored at –25°C.
11. The average diameter of emulsion is 20 μm. The DNA library is diluted with TE buffer to an average of 0.33 molecules per compartment. DNA concentration is determined by measuring absorbance at 260 nm. The aqueous phase consists of 1× *ExTaq* buffer, 0.2 mM of each dNTP, 0.25 μM M13M1cc, 0.25 U/μL of *ExTaq* DNA polymerase (Takara), random library DNA, 8×10^6 pBlue-Reverse-NH_2-coupled beads (see Note 7), and 2.5 nM pBRN, in a total volume of 20 μL. The oil phase is prepared by dissolving 1% Sun Soft No. 818SK (Taiyo Kagaku) in 380 μL of mineral oil (Sigma).
12. The aqueous mixture is added gradually to the mineral oil phase in a 5 mL vial while stirring at 750–800 rpm using a magnetic bar (10×4 mm), then the mixture is stirred for 3 min at room temperature (see Notes 8 and 9).
13. The obtained emulsion is dispensed in 50 μL increments into separate PCR tubes. Thereafter, emulsion PCR is conducted in the following temperature sequence: preheating at 95°C for 5 min; 40–45 cycles consisting of at 93°C for 15 s, 50°C for 30 s, and 72°C for 15 s, followed by an additional extension at 72°C for 7 min. The anneal temperature and the extension time should be rearranged as appropriate for your experiments (see Note 10).
14. After the PCR cycles, the emulsions are precipitated by centrifuging at 17,400×*g* for 3 min. The oil phase is then removed from the tubes to recover the emulsions including the reaction beads.
15. In order to disrupt the emulsions, 50 μL B&W buffer is added to each 0.5-mL tube. After mixing, the beads are precipitated by centrifuging at 17,400×*g* for 3 min. Then, the bead suspension is washed four times with 400 μL of hexane in a 0.5-mL tube, to completely remove the oil phase.
16. All the bead suspension is then transferred to a 1.5-mL tube and washed three times with 1-mL TE buffer. The beads are suspended in 10-μL TE buffer and stored at 4°C.

3.2. Preparation of Tagged Transcription Factor

1. DNA template for the expression of His-tagged transcription factor, PhaR, is amplified from pRSETphaR-His (6) using *ExTaq* DNA polymerase with 0.5 μM of each primer (F1 primer and R1 primer) in the following temperature sequence: preheating at 94°C for 5 min; 25 cycles consisting of at 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min, followed by an additional extension at 72°C for 7 min.
2. The PCR products are extracted using phenol–chloroform, precipitated with ethanol, and dissolved in TE buffer.
3. In vitro transcription/translation reactions are performed in 30 μL increments with components using S30 extract from a double protease-deficient mutant strain (Δ*clpP-ompT*) (7) for 1 h at 37°C.
4. After the reaction, the in vitro transcription factor-expression solution is stored at 4°C (see Notes 11 and 12).

3.3. High-Throughput Screening of GLOBE Using Flow Cytometry with Cell-Free Protein Synthesis

1. The DNA–bead complex solution (2 μL containing approximately 1.6×10^6 beads) is added to 10 μL of the in vitro expression solution and mixed by a rotator for 25 min at room temperature.
2. The beads are collected using a magnet and washed with 1 mL PBS (see Note 13). The collected beads are mixed with 100 ng of Anti-His(C-term)-FITC (Invitrogen) in 10 μL PBS by rotator at room temperature for 15 min under shaded conditions.
3. After the supernatant has been discarded (using a magnet), the beads are suspended in 500 μL PBS and the fluorescent events are sorted (using EPICS ELITE ESP [Beckman Coulter]). The bead complex fluoresces in green only when the His-tagged transcription factor interacts with DNA on the bead. Examples of the scattergram and histogram are shown in Fig. 2 (see Note 14).
4. The collected beads are suspended in TE buffer and stored at 4°C (see Note 15).

3.4. Recovery of DNA from GLOBE

1. DNA–bead conjugates are critically diluted with TE buffer, and dispensed to PCR tubes (an average of one bead per tube).
2. The templates are amplified separately in a 20 μL PCR mixture containing 0.025 U/μL *ExTaq* DNA polymerase (Takara) and 0.25 μM of each primer (M13M1cc and pBRN) in the following temperature sequence: preheating at 94°C for 5 min; 28 cycles consisting of at 94°C for 15 s, 50°C for 15 s, and 72°C for 15 s, followed by an additional extension at 72°C for 7 min.
3. The PCR products are checked on a 2.5% agarose gel (see Note 16).

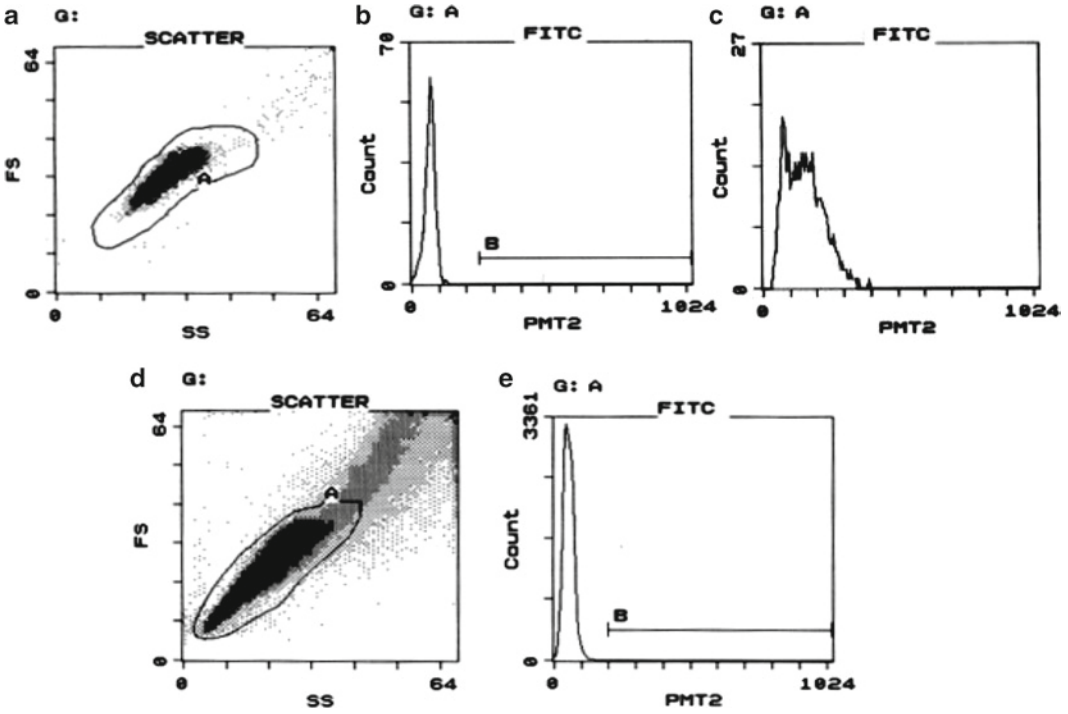


Fig. 2. Flow cytometry of beads complex. (**a**) Typical *dot-plot* of beads containing Pp (PhaR target sequences are involved) and negative-control fragment mixed with the in vitro-expressed transcription factor, PhaR, in FS (forward scatter) and SS (side scatter). Events gated in A (approximately 95% of total events) are subjected to the following analysis. (**b**) Typical histogram of "negative-control" beads (6) mixed with the in vitro-expressed transcription factor, PhaR, in FITC fluorescence intensity (for A-gated events). (**c**) Typical histogram of "positive-control" beads (6) mixed with the in vitro-expressed transcription factor, PhaR, in FITC fluorescence intensity (for A-gated events). (**d**) Typical *dot-plot* of GLOBE mixed with the in vitro-expressed transcription factor, PhaR, in FS and SS. Events gated in A (approximately 91% of total events) are subjected to the following analysis. (**e**) Typical histogram of GLOBE mixed with the in vitro-expressed transcription factor, PhaR, in FITC fluorescence intensity (for A-gated events). Events in region B are sorted. *PMT2* fluorescence channel 2 (FITC) signal intensity (reproduced from ref. 6).

4. The PCR products are extracted using phenol–chloroform, precipitated with ethanol, and dissolved in TE buffer.

3.5. Sequencing Analysis of DNA from GLOBE

1. The PCR products described in Subheading 3.4, step 4 are added to the ExoSAP reaction mixture in a total volume of 10 µL. After incubation at 37°C for 30 min, the reactions are stopped by incubation at 80°C for 15 min.
2. The DNA fragments are sequenced using the dye-terminator sequencing method with the pBlue-Reverse-bio primer.

3.6. Gel-Mobility Shift Assay Linked with In Vitro Transcription/ Translation

1. ExoSAP-treated PCR products described in Subheading 3.5, step 1 are extracted using phenol–chloroform and precipitated with ethanol.
2. After dilution with TE, the purified DNA templates are amplified separately in a 20 µL PCR mixture containing 0.05 U/µL

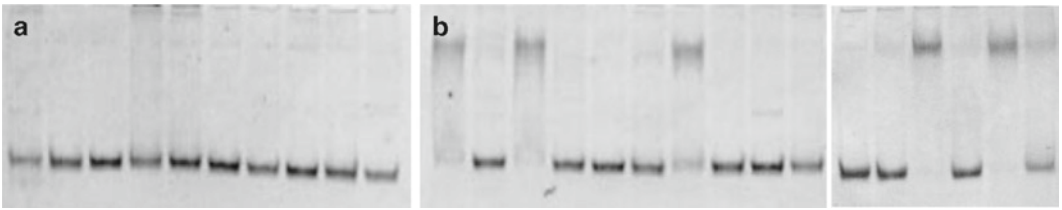


Fig. 3. Gel-mobility shift assay using Cy3-labeled DNA. Cy3-labeled DNAs are mixed with the in vitro-expressed transcription factor, PhaR, and gel-mobility shift assay is carried out. The Cy3-DNA fragments in the gel were detected with Typhoon 9400 (GE Healthcare). (**a**) Before sorting of GLOBE. (**b**) After sorting of GLOBE (reproduced from ref. 6).

Pyrobest DNA polymerase (Takara), and 0.25 μM of each primer (M13M1cc and pBRN-Cy3) in the following temperature sequence: preheating at 94°C for 5 min; 25 cycles at 94°C for 10 s, 60°C for 5 s, and 72°C for 5 s, followed by an additional extension at 72°C for 7 min.

3. The Cy3-DNA fragments are extracted using phenol–chloroform, precipitated with ethanol, and dissolved in TE.
4. One microliter of the Cy3-DNA fragment solution (approximately 50 ng/μL) is added to 9 μL of the in vitro transcription factor-expression solution. The solution is incubated for 20 min at 25°C, and a gel-mobility shift assay is carried out using 0.5× TBE under shaded conditions (see Note 17).
5. After electrophoresis, the Cy3-DNA fragments in the gel are detected using the Typhoon 9400 (GE Healthcare), according to the manufacturer's instructions. Only the fragments containing transcription factor-binding region are shifted by the addition of the transcription factor. Example images are shown in Fig. 3.

4. Notes

1. A micropipette tip with the end cut off is used to add an appropriate amount of the Sun Soft No. 818SK surfactant to the oil phase. The surfactant volume can be checked using microbalance. The oil phase should be incubated for 10 min at 50°C for distribution of the surfactant, followed by degassing using vacuum pump for 5 min to remove dissolved air for prevention of the disruption of the emulsion. In the recent study, an alternative oil phase (4% Sunsoft No. 818SK and 1% Span 80 in mineral oil) is used.
2. The transcription factor PhaR is derived from the methylotrophic bacterium *Paracoccus denitrificans* (8). The binding domain gene was placed under the T7 promoter to conduct

another transcription factor analysis. The construct is used as a template for cell-free protein synthesis.

3. *Escherichia coli* or other cell-free protein synthesis systems are commercially available from Roche Diagnostics, Invitrogen, or Shimadzu Biotech.
4. In this sequencing analysis, it should be noted that if the purity of sequencing primer is high, it is not always necessary to use the "biotinylated primer." The most important factor is sequencing primer purity in the analysis.
5. The EDC solution should be prepared freshly and added to the beads-primer solution as quickly as possible.
6. In order to ensure complete annealing, the temperature should be reduced to 30°C.
7. To remove the EDTA, the beads-primer solution should be washed with 100 μL sterile water (S.W.).
8. Alternatively, vortex mixing can also be used to make the emulsions. After gradually adding the aqueous mixture to the mineral oil phase in a 1.5-mL tube, the contents of the tube are thoroughly mixed with a vortex mixer for 15 s.
9. For checking the emulsions size, take a sample (~2 μL) from this emulsion and observe the emulsion state under a microscope.
10. In ref. 6, the emulsion PCR was carried out in the following temperature sequence: 55 cycles consisting of at 95°C for 15 s, 50°C for 9 min, and 72°C for 15 s, followed by an additional extension at 72°C for 7 min.
11. The in vitro expression solution should be used within a day.
12. An epitope-tagged transcription factor is prepared by using either the in vivo or in vitro protein synthesis system. The protein purification step is not indispensable. An example using cell-free protein synthesis is described in this chapter.
13. We have found that the "ice-cold" wash buffer effectively reduces the complex dissociation.
14. The sorting region setting is a critical factor for this screening. We have found that the sorting region under stringent conditions (B-region in Fig. 2) is more effective for screening.
15. The sorted beads–DNA complexes are unstable; thus, the next DNA amplification step should be carried out within a day.
16. Unless the amplification can detected, PCR cycles should be adjusted according to the number of template beads–DNA formed. In the case of irrelevant cycle setting, only smear bands will be detected.
17. For preventing the dissociation of the DNA–protein complex, SDS-free loading buffer should be used for the gel-mobility shift assay.

Acknowledgments

This work was supported in part by a Grant-in-Aid (No. 19360373) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

1. Pollack, J. R. and Iyer, V. R. (2002) Characterizing the physical genome. *Nat. Genet.* **32**, 515–21.
2. Orlando, V. (2000) Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem. Sci.* **25**, 99–104.
3. Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. L., Wilson, C. J., Bell, S. P., and Young, R. A. (2000) Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–9.
4. Roulet, E., Busso, S., Camargo, A. A., Simpson, A. J., Mermod, N., and Bucher, P. (2002) High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites. *Nat. Biotechnol.* **20**, 831–5.
5. Kojima, T., Takei, Y., Ohtsuka, M., Kawarasaki, Y., Yamane, T., and Nakano, H. (2005) PCR amplification from single DNA molecules on magnetic beads in emulsion: application for high-throughput screening of transcription factor targets. *Nucleic Acids Res.* **33**, e150.
6. Kojima, T., Yamane, T., and Nakano, H. (2006) In vitro selection of DNA binding sites for transcription factor, PhaR, from *Paracoccus denitrificans* using genetic library on microbeads and flow cytometry. *J. Biosci. Bioeng.* **101**, 440–4.
7. Ali, M., Suzuki, H., Fukuba, T., Jiang, X., Nakano, H., and Yamane, T. (2005) Improvements in the cell-free production of functional antibodies using cell extract from protease-deficient *Escherichia coli* mutant. *J. Biosci. Bioeng.* **99**, 181–6.
8. Maehara, A., Taguchi, S., Nishiyama, T., Yamane, T., and Doi, Y. (2002) A repressor protein, PhaR, regulates polyhydroxyalkanoate (PHA) synthesis via its direct interaction with PHA. *J. Bacteriol.* **184**, 3992–4002.

Chapter 23

PCR DNA-Array Profiling of DNA-Binding Transcription Factor Activities in Adult Mouse Tissues

Yimin Sun, Jing Cheng, and Keith R. Mitchelson

Abstract

Differential gene expression is tightly controlled by transcription factors (TFs), which bind close to target genes and interact together to activate and coregulate transcription. Bioinformatics analysis of published genome-wide gene expression data has allowed the development of comprehensive models of TFs likely to be active in particular tissues (signature TFs); however, the predicted activities of many of the TFs have not been experimentally confirmed. Here, we describe methods for the parallel analysis of the activities of more than 200 transcription factor proteins, using an advanced oligonucleotide array-based transcription factor assay (OATFA) platform, to assay TF activities in mice. The system uses a PCR-based system to translate cellular levels of target DNA–TF complex into a dye-tagged DNA signal, which is read by the developed microarray. The PCR step introduces semiquantitative amplification of the represented TF binding sequences. Experimental OATFA findings can identify many TF activities, which bioinformatics profiling does not predict. Newly identified TF activities can be confirmed by antibody-ELISA against active TFs. The PCR-based OATFA microarray analysis is a comprehensive method that can be used to reveal transcriptional systems and pathways which may function in different mammalian tissues and cells.

Key words: Transcription factor activities, Oligonucleotide array-based transcription factor assay, Single primer amplification

1. Introduction

Studies in mammalian tissues show that transcriptional regulation of genes of related function is highly regulated and that coordinated transcription occurs in genes related to particular tissue-specific functions (1–3). Experimental studies of tissue-specific gene transcription suggest that TFs act via *regulatory TF modules* to confer tissue specificity in which TFs are brought in close (physical) contact by binding to clusters of adjacent TF binding sites (4). The modular concept proposes that several bound

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_23, © Springer Science+Business Media, LLC 2011

coincident TFs interact to cause the activation of a particular gene. The large numbers of genes and their many potentially interacting TFs make the task of experimental identification of the activity of any particular TF in any tissue or cell type both extremely time-consuming and technically difficult. Thus currently, experimental knowledge of tissue-specific TF activities is limited.

1.1. OATFA Platform

Given that mammalian TFs regulate transcription mainly through their DNA-binding activities, we developed methods for high-throughput quantitative assessment of protein–DNA-binding specificities which permit a relative measure of TF protein activities. We initially profiled human TF activity by an oligonucleotide array-based transcription factor assay (OATFA) (5). OATFA uses an *in-solution* approach to simultaneously analyze the DNA-binding activities of TFs, and the binding activities are then resolved indirectly by use of an oligonucleotide array-based assay of the associated DNA rather than by detecting the bound transcription factor. The solution-based assay approach obviates potential surface-associated complications (6). We improved the original OATFA method by the introduction of single primer (PCR) amplification (SPA) which increases assay sensitivity and allows semiquantitative estimation of the relative TF activities and reported the development of an advanced human OATFA platform (7). Here, we describe the application of an OATFA platform specifically for mouse (MOUSE OATFA) (see Fig. 1) which can rapidly analyze some 200 TF activities simultaneously, representing about 10% of the predicted murine TFs (8). Highly selective probe design ensures high specificity of TF capture, while the signal output permits the semiquantitative representation of TF activities. The use of SPA was critical for providing sufficient sensitivity to permit semiquantitative estimations of TF activity by OATFA, and data can then be compared with other supposedly quantitative methods, such as relative TF activities inferred from TF gene expression measurements, TF protein levels and antibody-based ELISA measurements of active TF activities, etc.

1.2. TF Activities in Adult Mouse Tissues

Here, a comparative analysis of the intrinsic activities of DNA-binding TFs is made across a panel of six adult mouse tissues, liver, kidney, lung, brain, heart, and skeletal muscle (Fig. 2a), to illustrate the method. Several observations can be made from the findings:

1. The hierarchical clustering of the data confirms that tissue from different individual animals with similar types of function tend to cluster together. For example, the tissue pairs of heart and skeletal muscle (muscle), liver and kidney formed two cluster pairs because of the similarity of many biological processes and functions which are reflected in the relative TF activities found in the tissues. These observations are similar

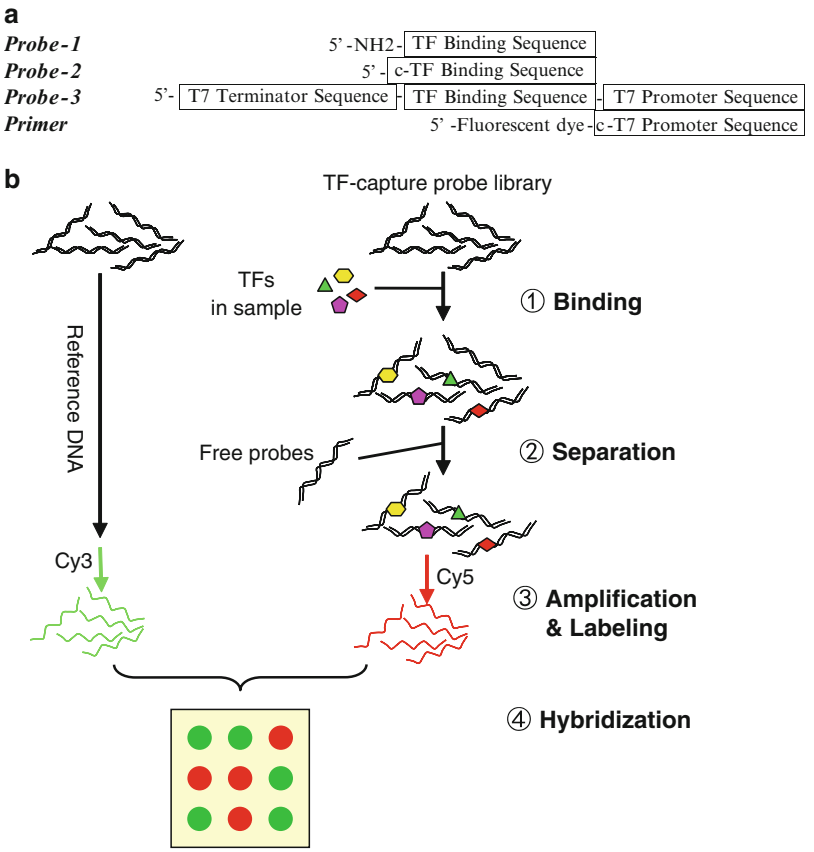


Fig. 1. The MOUSE OATFA single sample assay (SSA) procedure. (**a**) Probe sets. *Boxed areas* are nucleotide sequences: *Probe-1*: microarray capture probe; *Probe-2* and *Probe-3*: TF capture duplex DNA; *Primer*: Fluorescent-labeled signal amplification primer. (**b**) The SSA assay steps are: (1) binding interactions between tissue TFs and their specific duplex DNA-binding probes; (2) electrophoretic separation of bound and free probes; (3) amplification and labeling of (purified TF captured) DNA probes; (4) hybridization of probes to the MOUSE OATFA microarray. *Reference DNA*: The TF capture DNA library was directly amplified and counterlabeled as the reference DNA to be hybridized with labeled sample DNA. Reprinted from (8) with permission. Copyright 2008 American Chemical Society.

to the conclusions of a cluster analysis of global mRNA expression activities across various mouse tissues (11).

2. The TFs could be grouped by their relative activities in the 6 examined tissues into three general categories (Fig. 2b): being either active in each of the tissues (*Cluster 1*) or not active in any of these tissues (*Cluster 2*) or active in some tissues (*Cluster 3*). Tissue-specific TF activity (*Cluster 3*) is an area of intense research interest, and numerous bioinformatics-based predictions of tissue-specific TF interactions (12–15) have been inferred from the readily measurable tissue-specific gene expression data of prospective TF targets and from knowledge of the TF binding sequences present in gene promoter and adjacent regulatory regions. In our previous report (8) on MOUSE OATFA, we compared the findings between

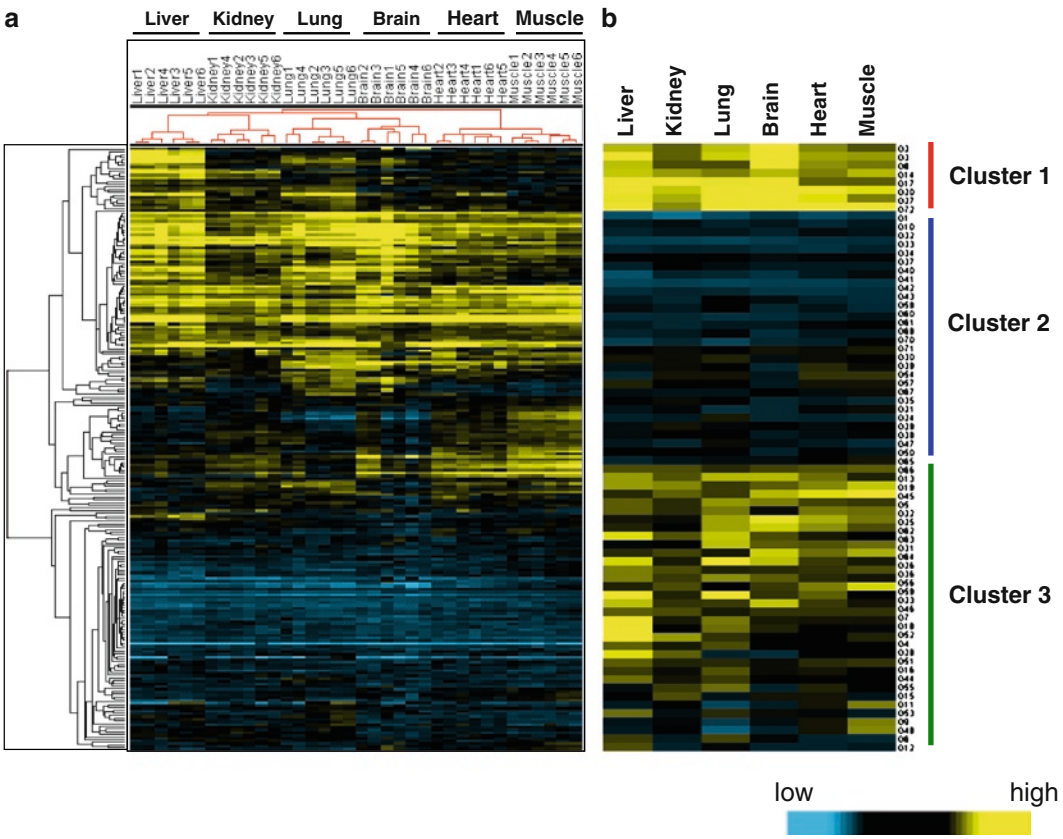


Fig. 2. The hierarchical clustering of the MOUSE OATFA signals of the activities of 200 different TFs in six mouse tissues. (**a**) The entire data set comprising 6 replicates of each tissue. (**b**) Selected representative MOUSE OATFA signals of subsets of TFs. *Cluster 1*: The TFs were active in each tissue. *Cluster 2*: The TFs were not detected in each of the tissues (*below detection limits*). *Cluster 3*: The TFs were active in some of the six tissues. Reprinted from (8) with permission. Copyright 2008 American Chemical Society.

OATFA and four other published predictions of tissue-specific TF activities in each of the six tissues. We also reported correlations of the semiquantitative levels at which TFs are detected by MOUSE OATFA in tissues with the semiquantitative levels at which the TF gene itself is expressed in those tissues as determined by Su et al. (16). While there is general high agreement between the methods (see Note 1), there are also significant differences in the measured activities which we attempted to resolve using independent quantitative ELISA assays of active TFs.

1.3. Correlation of MOUSE OATFA Tissue-Specific TF Activity (Cluster 3) with Gene Expression Activities

The active TFs detected by MOUSE OATFA in specific tissues show close agreement (77% concordance) with bioinformatics predictions of tissue-specific TF activities based on relative gene expression of TF gene targets in different tissues (12–15). We suggest that the direct assay of TF activity by the MOUSE OATFA system could provide an independent validation of the predicted

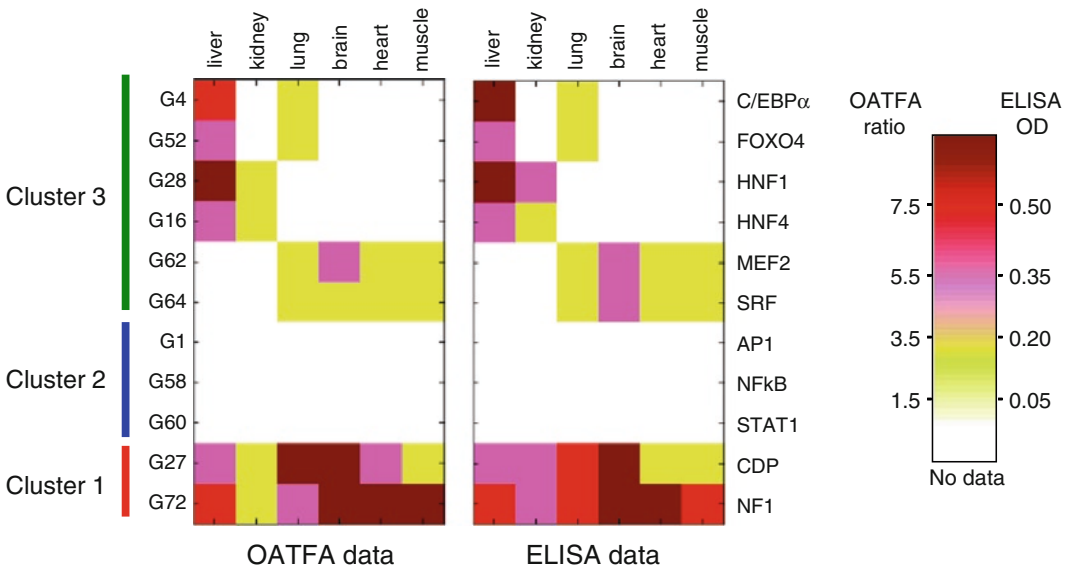


Fig. 3. Comparison of MOUSE OATFA signals of 11 TFs with their quantitative ELISA estimations in six mouse tissues. Representative subsets of TFs were indicated (*left side*). *Cluster 1*: The TFs were active in each tissue. *Cluster 2*: The TF activities were not detected in each of the tissues (*below detection limits*). *Cluster 3*: The TFs were active in some of the six tissues. The name code for individual TFs are indicated (*right side*). Reprinted from (8) with permission. Copyright 2008 American Chemical Society.

tissue-specific TF activities inferred from bioinformatics analysis. However, important differences between MOUSE OATFA and bioinformatics analyses can also be seen. For example, the transcription factor SRF is detected by MOUSE OATFA in mouse brain and lung, but SRF activity is not inferred in these tissues by bioinformatic analysis, since the expression of the TF gene itself has not been reported (Fig. 3). Such differences in TF activity detected rapidly by MOUSE OATFA are entirely verifiable by independent methods such as TF-ELISA (see Subheading 3.3). The concordance between MOUSE OATFA and TF-ELISA assays indicates the power of MOUSE OATFA to identify new TF activities in tissues and organs.

2. Materials

2.1. Reagents

1. All oligonucleotide probes (desalted) are from Operon Technologies Inc.
2. Capture *Probe-1* oligonucleotides are diluted in Spotting Solution (CapitalBio Corp) for microarray slide printing.
3. Antibodies for AP1 (#sc-1694x), CDP (#sc-6327), C/EBPα (#sc-61), FOXO4 (#sc-34903), HNF1 (#sc-8986), HNF4

(#sc-8987), MEF2 (#sc-313x), NF1 (#sc-5567), NFκB (#sc-372x), SRF (#sc-335), and STAT1 (#sc-346) are from Santa Cruz Biotechnology Inc.

4. NE-PER® Nuclear and Cytoplasmic Extraction Reagents, chemiluminescent SuperSignal® reagent and BCA™ Protein Assay Kit are from Pierce.
5. CER I and CER II solutions, protease inhibitor, and nucleotide triphosphates are each from Roche Applied Science.
6. QIAEX II® Gel Extraction Kit is from Qiagen Inc.
7. 1× PCR buffer (Mg^{2+} plus) and TaKaRa *Taq*™ are from TaKaRa.
8. All general buffers and biochemicals are of molecular biology grade, and the agaroses and general chemicals are of A.R. grade, all from Sigma Inc.
9. The 1× TF binding buffer contains 10 mM Tris–HCl, pH 7.5, with 50 mM NaCl, 0.5 mM EDTA, 1 mM $MgCl_2$, 4% glycerol, 0.5 mM dithiothreithol, 0.05 μg/μl poly (dI-dC) (see Note 2). Poly (dI-dC) is from Amersham Biosciences.
10. 3× SSC (450 mM sodium chloride/45 mM sodium citrate, pH 7.0).
11. Loading buffer (0.25× TBE, 60%; glycerol, 40%).

2.2. OATFA Probe Design

The oligonucleotide probes are based on TRANSFAC® Professional r8.2 (http://www.biobase.de), a database of eukaryotic transcription factors, their genomic binding sites, and DNA-binding profiles (see Note 3). Fuller details of the considerations for the design of the transcription factor binding probes are given in Qiao et al. (7) and Sun et al. (8).

2.3. Probe Sets

1. Each "TF-probe set" comprises three subsidiary oligonucleotide probes: *probe-1*, *probe-2*, and *probe-3* (Fig. 1a). *Probe-1* is bound by 5′-amino group to the surface of the glass slide and serves to capture (hybridize) probe-2 of each set. *Probe-2* has the reverse complementary TF binding sequence to the sequences in *probe-1* and *probe-3*. *Probe-3* contains an additional T7 promoter sequence (5′-CCCTATAGTGAGTCGTATTACCCC-3′) at the 3′-terminus (see Note 4).
2. The annealed products of *probe-2* and *probe-3* of each set form duplex regions that are bound by a particular transcription factor. Four probe sets for the control of TF-DNA binding (negative control for binding, BNC) ensure the reliability of microarray analysis. The signal intensities of the BNCs reveal the levels of nonspecific binding in a MOUSE OATFA microarray assay. Four probes (*probe-1*) for control of hybridization (negative control for hybridization, HNC) are also

designed with low similarity to all other position-specific scoring matrices (PSSMs). HNC probes are used as positive controls for microarray printing process. The "Total TF-probe library" mixture is composed of the 240 duplex TF-capture probes and the four BNC probes (see Note 5).

3. Methods

3.1. Preparation of Mouse Nuclear Extracts

1. In the illustrated method, adult (10- to 12-week-old) mouse tissues are dissected from male ICR mice and snap-frozen in liquid nitrogen and stored at −80°C until use.
2. Nuclear extracts are prepared using NE-PER® Nuclear and Cytoplasmic Extraction Reagents (Pierce, Rockford, IL) following the manufacturer's instructions.
3. Homogenized tissues are incubated on ice in CER I solution with protease inhibitor for 10 min.
4. CER II reagent is added, and after incubation for an additional 1 min on ice, nuclei are collected by centrifugation at 16,000 × *g* for 5 min.
5. Nuclei are then lysed on ice in NER buffer containing protease inhibitor for 40 min with intense vortexing every 10 min.
6. Centrifugation at 16,000 × *g* for 10 min at 4°C is used to clear cellular debris from the soluble nuclear extract (NE), which is then dispensed into aliquots and stored at −80°C until use (see Note 6).
7. Protein concentration of the NE preparations is determined using BCA™ Protein Assay Kit.

3.2. OATFA Analysis of TF Activities

The method for MOUSE OATFA uses the single sample assay (SSA) format (Fig. 1b) in which TF signals are corrected for residual unbound capture probe signals. Other analysis formats can also be used for OATFA analysis, see Qiao et al. (7). The MOUSE OATFA platform described here employs five basic steps:

1. *Binding.* A set of duplex DNA-binding oligonucleotides (TF-probe mixture) are incubated with a nuclear extract to allow the formation of DNA/protein complexes. TFs and duplex probes are allowed to interact in a 20-µl mixture containing 1× binding buffer, purified TFs, or cell nuclear extracts at room temperature for 60 min. The nuclear extracts (10–15 µg) are then incubated with the binding buffers for 10 min on ice to reduce possible nonspecific binding, prior to the addition of the double-stranded TF-probe mixture in

which each TF-capture probe is present at a final concentration of 0.1 nM.

2. *Separation*. The DNA/protein complexes are separated from any residual unbound DNA probes by agarose gel electrophoresis. After 1-h incubation at 30°C, the 20-µl reaction mixture is mixed with 2 µl of loading buffer (0.25× TBE, 60%; glycerol, 40%) and is then loaded into separate lanes on 1.5% agarose or 2% agarose gels (Sigma), and electrophoresis is performed in chilled 0.5× TBE buffer for 30 min at 12 V/cm to separate DNA–TF protein complexes away from unbound DNA probes. The gel area containing the protein/DNA complex is excised and transferred into a 1.5-ml tube. Gel extraction is performed by QIAEX II® Gel Extraction Kit according to the manufacturer's instructions.
3. *Amplification and labeling*. The DNA probes are then extracted from the TF protein complexes and the DNA quantitatively SPA (PCR) amplified using a 5′-fluoro-tagged single strand primer. Cy3- and Cy5-labeled T7 promoter sequence (5′-GGG GTA ATA CGA CTC ACT ATA GGG-3′) is used as the primer oligonucleotide (see Fig. 1a) for single-primer PCR amplification (9). The 20-µl reaction mixture contained a mixture of recovered double-stranded DNA probes as the template(s), 1× PCR buffer (Mg^{2+} plus), 200 nM Cy3- or Cy5-labeled primers, 200 µM mix of four dNTPs, and 1 U TaKaRa *Taq*. The SPA amplification involves an initial denaturation step at 95°C for 5 min; 30 cycles of denaturation at 95°C for 30 s, annealing at 53°C for 30 s, and elongation at 72°C for 20 s and then a final single extension step at 72°C for 10 min. The SPA products are heated to 45°C and then vacuum-dried.
4. *Hybridization*. The probes are hybridized to the profiling microarray and provide a semiquantitative estimation of relative TF activities via the estimation of the amount of probe DNA captured by each TF. The two corresponding SPA products with either Cy3- or Cy5-label are mixed and hybridized to the MOUSE OATFA microarrays (see Note 7) in 3× SSC, 0.2% SDS, 5× Denhardt's solution, and 25% formamide in a total 12-µl mixture at 42°C for 16–20 h. The arrays are then washed sequentially at 42°C in 2× SSC plus 0.5% SDS for 10 min and then 0.2× SSC plus 0.1% SDS for 10 min. Immediately after washing, the array slides are spun dry in a centrifuge at 200×*g* for 2 min prior to scanning.
5. *Scanning and data analysis*. Array images are acquired by a laser confocal scanner LuxScan™-10KA, and signal intensities for each spot were calculated by subtracting local background using LuxScan 3.0 software (both CapitalBio Corp). The

DNA-binding activity of a TF is regarded as significant if the median ratio is >1.5 and P value is <0.05 (8). Hierarchical clustering using an average linkage algorithm is performed with an output of dendrograms generated by TreeView (10).

3.3. Independent Measure of TF Activity by Antibody-Based ELISA Analysis

Selected tissue-specific TF activities can be independently assayed using antibody-based ELISA (see Note 8). Here, 11 transcription factors assayed by TF-ELISA are illustrated, including a number of previously known as well as novel TF activities: Two TFs from *Category 1* (NF1, CDP), three TFs from *Category 2* (STAT1, NFkB, AP1), and six TFs from *Category 3* (SRF, MEF2, HNF4, HNF1, FOXO4, C/EBPα). The quantitative TF-ELISA analysis of each of the 11 TFs agree closely with semiquantitative estimates of their activity determined by MOUSE OATFA (Fig. 3). HNF1 and HNF4 are members of a hepatocyte nuclear factor family known to regulate tissue-specific gene expression in liver and kidney (13) which are confirmed in MOUSE OATFA and ELISA assays. FOXO4 (AFX) is a member of the forkhead FOXO transcription factor family which Yang and colleagues (17) predicted could participate in the regulation of eye and liver tissues despite their lack of direct experimental data, by inference from its known interaction with PBX1A and FOXO1A which are expressed in those respective tissues. MOUSE OATFA (8) shows that FOXO4 has highest activity in adult mouse liver and next highest in lung which is also confirmed by ELISA assays (Fig. 3).

Several TFs representative of the G63, G62, G5, and G64 classes which here correspond to the RSRFC4, MEF2, MyoD family, and SRF are here found by MOUSE OATFA to have similar relative activities in lung, brain, heart, and skeletal muscle (Fig. 2). Thus, in addition to their known presence in heart and muscle, MOUSE OATFA detects the four TFs in lung and brain, previously unreported (8). These MEF2 and SRF activities are also verified by the ELISA assay (Fig. 3). Yu and colleagues (13) predicted that RSRFC4, MyoD, and SRF each interact with MEF2, while others have reported that the four TFs are involved in the development of skeletal muscle and heart (1, 18).

3.4. Contrasting the Measured TF Activities

Here, the MOUSE OATFA analysis indicates that the highest activity of C/EBPα is in liver, followed by lung, while the other four tissues display no detectable activity, and each result is confirmed quantitatively by independent ELISA analysis (Fig. 3). ELISA estimates of the SRF activity also agree closely with MOUSE OATFA determinations, and both methods showed the highest activity of this TF in brain, followed by roughly equivalent high activity in lung, heart, and muscle (Fig. 3). Although several other assay methods (e.g., Western blotting, mRNA expression of TF genes) are also frequently used to indicate potential TF activity, our findings suggest that two methods that directly

assay active TF in cells – antibody-based ELISA and OATFA assays – more usually agree, and we suggest that these methods most likely reflect true TF activities.

3.5. Application of MOUSE OATFA

It is notable that although the current MOUSE OATFA platform measures only the activity of some 10% of the predicted mouse TFs, the tissue-specific signatures of TF activities can provide novel support for prediction of functional interactions between several TFs (see Note 9). Active TFs detected by MOUSE OATFA in specific tissues correspond closely (77% concordance) with bioinformatics-based predictions of tissue-specific TF activities (12–15) based on analysis of expression of TF gene targets. Some 86% of the active TFs detected by MOUSE OATFA in particular tissues are also confirmed by combined reports of the expression of the TF gene or by TF protein levels in those tissues (see Note 10). In addition to the agreement with previous reports, tissue-specific TF activities may also be predicted by MOUSE OATFA in which expression of the TF gene has not been reported previously (see Note 11).

Our findings suggest that systematic analysis of active TF functionality should focus on relative binding activities to gene promoter regulatory elements, rather than on the mRNA levels or the quantities of protein, because many TFs are only active following posttranslational protein modifications and are not reflected in absolute protein levels. We suggest that ideally two types of genome-wide information are required for bioinformatics-based development of network models and active tissue-specific pathways in mouse (see Note 12). Gene expression data may be used to identify global transcription (changes), and parallel analyses of TF activities by MOUSE OATFA can provide guidance to TF activity predictions (see Note 13).

4. Notes

1. There is a high concordance (86%) between the specific tissue in which TF gene expression or the TF protein has been independently reported and the active TF in that tissue seen by MOUSE OATFA assay (8).
2. As used commonly in the conventional EMSA (19), specific TF capture–probe interactions are enhanced by the addition of poly dI-dC, which reduces nonspecific competition by other proteins that can compete with TFs for binding to their specific TF binding sequences.
3. An activity-based probe design method based on a simulated annealing algorithm (20, 21) is used to profile the large number of transcription factors for score optimization and for the estimation of the hybridization parameters of each probe,

including the melting temperature (T_m), secondary structure formation, dimer formation, cross hybridization, and sequence specificity. All probes were designed based on TRANSFAC® Professional r8.2 (http://www.biobase.de), a database of eukaryotic transcription factors, their genomic binding sites and DNA-binding profiles (22, 23). The principles of probe design include: (a) the highest sequence similarity with the target PSSMs (≥*minFP*) representing nucleotide distribution matrices for *cis*-acting DNA regulatory regions in the TRANSFAC® database, (b) the lowest sequence similarity with other PSSMs (≤*minFN*), and (c) the highest specificity of oligonucleotide hybridization.

4. Different PSSM matrix mammalian TF binding sequences (from TRANSFAC database) may have lengths ranging from 6 to 30 nt. In order to equalize microarray hybridization characteristics, the "binding region" of *probe-1* and *probe-3* oligonucleotides are each made same length (30 nt) by the addition of random nucleotides to the terminals of the matrix sequences less than 30-nt length.
5. Detailed descriptions of the design of probes and the sequences of all *human* OATFA probes are given in Qiao et al. (7), and all *mouse* OATFA probes are given in Sun et al. (8).
6. If cell cultures are used (rather than tissues), nuclear extracts can be prepared using the Nuclear Extract Kit (Active Motif, USA) according to the manufacturer's instructions.
7. *OATFA microarray fabrication*. The 5′-amine linked *Probe-1* oligonucleotides are diluted in Spotting Solution to 20 μM and then printed in duplicate in arrays of 22 rows × 24 columns under 45% humidity on aldehyde-modified glass slides and then coupled covalently overnight in 70% humidity (8).
8. Independent validations of single TFs can be used to confirm some of the MOUSE OATFA observations. Colorimetric microwell ELISA TF assays performed as described by Shao *et al.* (5) using antibodies against active forms of AP1, CDP, C/EBPα, FOXO4, HNF1, HNF4, MEF2, NF1, NFκB, SRF, and STAT1 could confirm the TF activities detected by the multiplex MOUSE OATFA assay.
9. In respect to data points generated, there is a large difference in the scale of analysis between gene expression microarrays (10,000 genes) compared to the MOUSE OATFA platform (200 TFs). Despite the differences in scale, the MOUSE OATFA platform provided additional, unique, and verifiable identifications of active transcription factors that are not seen by expression profiling. It is to be anticipated that further confirmations of known or predicted TF activities and further discovery of new tissue-specific TF activities will be made if the number of TFs assayed by the MOUSE OATFA platform is increased.

10. Individual TFs are typically found to be active in several tissues and only rarely in one tissue. Considering that only six mouse tissues were examined, it is to be expected that the majority of cellular TFs would be found active in more than one tissue in mouse. In different tissues, they interact with different coregulatory partners to coordinate the expression of different tissue-specific target genes.
11. The detection (see Fig. 3) of the activity of FOXO4 in liver and lung, of AR, ER, and IRF1 in muscle, of SRF in brain and lung, and of CDP in each of the surveyed tissues are examples of newly reported observations (8).
12. Sun et al. (8) reported the use of oPOSSUM software (12, 24), employing algorithms found at http://www.cisreg.ca, to develop a model of novel networks in mouse tissues. Some protein-on-protein interactions between particular TFs involved in these predicted networks has been reported previous to the study; however, the analysis by Sun et al. also identified numerous potential TF–TF interactions that could be tested. We suggest that MOUSE OATFA has important predictive power for directing the testing of potential TF–TF interactions in particular tissues.
13. The MOUSE OATFA array data here agrees closely with the known biological data from each tissue across the full range of surveyed TFs, with high levels of agreement with reported tissue-specific TF activities and good general agreement with many bioinformatically predicted tissue-specific TF activities. Our data further suggests that the power of the MOUSE OATFA platform is to provide unique information concerning the presence of active TFs in individual tissues, as inspection of the expression of the TF-encoding genes does not always provide an accurate representation of their likely activities.

Acknowledgments

This work was supported by the grant of the National Hi-Tech Program of China (No. 2006AA020701).

References

1. Wasserman, W.W. and Fickett, J.W. (1998) Identification of regulatory regions which confer muscle-specific gene expression. *J Mol Biol* **278,** 167–81.
2. Thompson, W., Palumbo, M.J., Wasserman, W.W., Liu, J.S., and Lawrence, C.E. (2004) Decoding human regulatory circuits. *Genome Res* **14,** 1967–74.
3. Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., Chen, L., Chen, L., Chen, T.M., Chin, M.C.,

et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445,** 168–76.

4. Krivan, W. and Wasserman, W.W. (2001) A predictive model for regulatory sequences directing liver-specific transcription. *Genome Res* **11,** 1559–66.
5. Shao, W., Wei, H.J., Qiao, J.Y., Zhao, Y.C., Sun, Y.M., Zhou, Y.X., and Cheng, J. (2005) Parallel profiling of active transcription factors using an oligonucleotide array-based transcription factor assay (OATFA). *J Proteome Res* **4,** 1451–6.
6. Egener, T., Roulet, E., Zehnder, M., Bucher, P., and Mermod, N. (2005) Proof of concept for microarray-based detection of DNA-binding oncogenes in cell extracts. *Nucleic Acids Res* **33,** e79.
7. Qiao, J.Y., Shao, W., Wei, H.J., Sun, Y.M., Zhao, Y.C., Zhang, L., Mitchelson, K., and Cheng, J. (2008) Novel high-throughput profiling of human transcription factors and its use for systematic pathway mapping. *J Proteome Res* **7,** 2769–79.
8. Sun, Y.M., Zhang, Y., Zeng, L.Q., Wu, J.P., Wei, L., Ren, A.H., Shao, W., Qiao, J.Y., Zhao, Y.C., Zhang, L., Mitchelson, K.R., and Cheng, J. (2008) Broad profiling of DNA-binding transcription factor activities improves regulatory network construction in adult mouse tissues. *J Proteome Res* **7,** 4455–64.
9. Smith, L., Underhill, P., Pritchard, C., Tymowska-Lalanne, Z., Abdul-Hussein, S., Hilton, H., Winchester, L., Williams, D., Freeman, T., Webb, S., and Greenfield, A. (2003) Single primer amplification (SPA) of cDNA for microarray expression analysis. *Nucleic Acids Res* **31,** e9.
10. Saldanha, A.J. (2004) Java Treeview-extensible visualization of microarray data. *Bioinformatics* **20,** 3246–8.
11. Zhang, W., Morris, Q.D., Chang, R., Shai, O., Bakowski, M.A., Mitsakakis, N., Mohammad, N., Robinson, M.D., Zirngibl, R., Somogyi, E., Laurin, N., Eftekharpour, E., Sat, E., Grigull, J., Pan, Q., Peng, W.T., Krogan, N., Greenblatt, J., Fehlings, M., Kooy, D., Aubin, J., Bruneau, B.G., Rossant, J., Blencowe, B.J., Frey, B.J., and Hughes, T.R. (2004) The functional landscape of mouse gene expression. *J Biol* **3,** 21.
12. Ho Sui, S.J., Mortimer, J.R., Arenillas, D.J., Brumm, J., Walsh, C.J., Kennedy, B.P., and Wasserman, W.W. (2005) oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic Acids Res* **33,** 3154–64.
13. Yu, X., Lin, J., Zack, D.J., and Qian, J. (2006) Computational analysis of tissue-specific combinatorial gene regulation: predicting interaction between transcription factors in human tissues. *Nucleic Acids Res* **34,** 4925–36.
14. Smith, A.D., Sumazin, P., and Zhang, M.Q. (2007) Tissue-specific regulatory elements in mammalian promoters. *Mol Syst Biol* **3,** 73.
15. Barrera, L.O., Li, Z., Smith, A.D., Arden, K.C., Cavenee, W.K., Zhang, M.Q., Green, R.D., and Ren, B. (2008) Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. *Genome Res* **18,** 46–59.
16. Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M.P., Walker, J.R., and Hogenesch, J.B. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* **101,** 6062–7.
17. Yang, H.L., Zhao, R.Y., Yang, H.Y., and Lee, M.H. (2005) Constitutively active FOXO4 inhibits Akt activity, regulates p21 Kip1 stability, and suppresses HER2-mediated tumorigenicity. *Oncogene* **24,** 1924–35.
18. Weintraub, H., Dwarki, V.J., Verma, I., Davis, R., Hollenberg, S., Snider, L., Lassar, A., and Tapscott, S.J. (1991) Muscle-specific transcriptional activation by MyoD. *Genes Dev* **5,** 1377–86.
19. Kolkman, A., Slijper, M., and Heck, A.J. (2005) Development and application of proteomics technologies in *Saccharomyces cerevisiae*. *Trends Biotechnol* **23,** 598–604.
20. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953) Equation of state calculations by fast computing machines. *J Chem Phys* **21,** 1087–92.
21. Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P. (1983) Optimization by simulated annealing. *Science* **220,** 671–80.
22. Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Pru, M., Reuter, I., and Schacherer, F. (2000) TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res* **28,** 316–9.
23. Matys, V., Fricke, E., Geffers, R., Gössling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V., et al. (2003) TRANSFAC®: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* **31,** 374–8.
24. Ho Sui, S.J., Fulton, D.L., Arenillas, D.J., Kwon, A.T., and Wasserman, W.W. (2007) oPOSSUM: integrated tools for analysis of regulatory motif over-representation. *Nucleic Acids Res* **35** (Web Server issue), W245–52.

Chapter 24

Nucleotide Exchange and Excision Technology DNA Shuffling and Directed Evolution

Janina Speck, Sabine C. Stebel, Katja M. Arndt, and Kristian M. Müller

Abstract

Remarkable success in optimizing complex properties within DNA and proteins has been achieved by directed evolution. In contrast to various random mutagenesis methods and high-throughput selection methods, the number of available DNA shuffling procedures is limited, and protocols are often difficult to adjust. The strength of the *n*ucleotide *ex*change and *ex*cision *t*echnology (NExT) DNA shuffling described here is the robust, efficient, and easily controllable DNA fragmentation step based on random incorporation of the so-called 'exchange nucleotides' by PCR. The exchange nucleotides are removed enzymatically, followed by chemical cleavage of the DNA backbone. The oligonucleotide pool is reassembled into full-length genes by internal primer extension, and the recombined gene library is amplified by standard PCR. The technique has been demonstrated by shuffling a defined gene library of chloramphenicol acetyltransferase variants using uridine as fragmentation defining exchange nucleotide. Substituting 33% of the dTTP with dUTP in the incorporation PCR resulted in shuffled clones with an average parental fragment size of 86 bases and revealed a mutation rate of only 0.1%. Additionally, a computer program (NExTProg) has been developed that predicts the fragment size distribution depending on the relative amount of the exchange nucleotide.

Key words: DNA shuffling, NExT, DNA recombination, Directed evolution, In vitro evolution, Protein engineering

1. Introduction

Nature employs mutation, selection, and recombination to evolve highly adapted individuals. Researchers can mimic these natural evolutionary processes by directed evolution techniques, searching through constructed libraries with billions of members for specific functions, for example, by ribosome display, phage display, or protein-fragment complementation assay (1–4). Several strategies to create randomly mutagenized gene libraries have

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4_24, © Springer Science+Business Media, LLC 2011

been described (5), of which error-prone PCR is the most commonly used. Additionally, site-directed, site-saturation mutagenesis, or DNA synthesis can be used to introduce localized diversity at the genetic level. After diversification, the library is screened or selected for specific traits of the encoded gene product. One of the most effective strategies in directed protein evolution is to gradually accumulate mutations, either sequentially or by recombination, while applying increasing selection pressure.

Most of the initially introduced mutations will not improve the protein with regard to the desired function. Selection and/or screening for improved function will lead to relative enrichment of advantageous mutations while removing the most disadvantageous ones from the library. However, if more than one mutation per gene is randomly introduced, most genes will contain disadvantageous mutations additionally to advantageous ones, even after repeated selection cycles. In vitro DNA recombination techniques combining gene fragments from previously enriched genes provide the possibility to outcross undesired and enrich desired mutations in newly mixed mutational patterns. Such recombined libraries can then be screened or selected for variants containing only advantageous mutations.

Various methods are available for in vitro recombination of genes. In the 'original' DNA shuffling protocol of Stemmer, DNA is fragmented by DNase digest, and fragments of the desired size are extracted from an agarose gel (6, 7). The purified fragments are reassembled with an internal primer extension reaction and amplified by standard polymerase chain reaction. However, the amount of nuclease as well as the digest conditions require careful optimization to generate fragments of desired length and to avoid too much DNA loss. Limited controllability of fragment size and recombination rate is also the main drawback of various other methods, such as the staggered extension process (StEP) (8) and random-priming recombination (RPR) (9).

In contrast, the nucleotide exchange and excision technology (NExT) DNA shuffling is easy to control and predictable (Fig. 1). The NExT procedure was originally developed using test libraries with three to six truncated mutants of chloramphenicol acetyl transferase I (CAT, a bacterial enzyme that mediates resistance to chloramphenicol) and dUTP as fragmentation-defining exchange nucleotide. In the mean time, the technique was applied to recombine several genes of different lengths using dUTP fractions of 25–35%. The amount of dUTP was chosen based on the fragment size prediction with our computer program NExTProg (10, 11). Without any other adjustment, intermediate tests or analysis steps, fragmentation and reassembly worked in the first attempt, even when recombining genes of only 284 bp length (data not shown).

Due to the robustness and simplicity of NExT DNA shuffling, even those with little previous experience in this area should be able to apply this technique for shuffling short genes or large gene

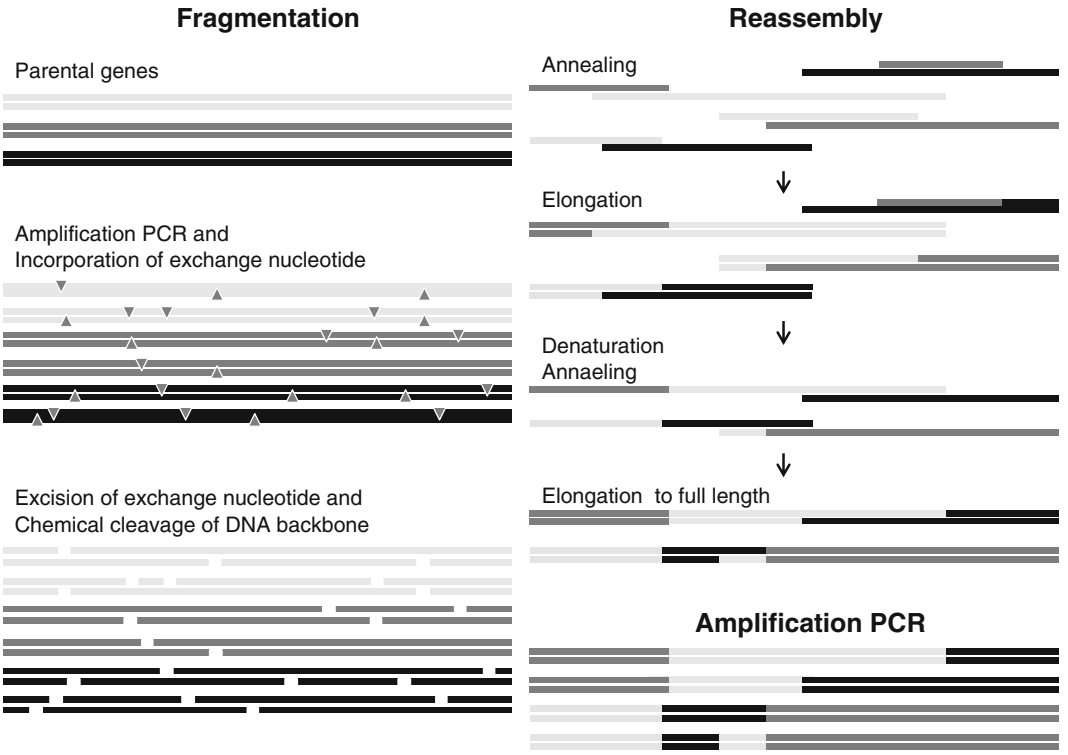


Fig. 1. Schematic overview of the nucleotide exchange and excision technology (NExT) DNA shuffling. *Fragmentation:* In the first step, a pool of homologous genes is amplified by a standard PCR in which a fragmentation-defining exchange nucleotide (*triangle*) is incorporated together with the four standard nucleotides. The exchange nucleotide is excised enzymatically, leaving an abasic site. Finally, the DNA is fragmented by chemical cleavage of the DNA backbone at the abasic sites. *Reassembly:* In this PCR-like reaction, no external primers are added. Overlapping fragments anneal with each other, serving as primer and template at the same time. With each round of denaturation, annealing, and elongation, the fragments are combined to DNA segments of increasing lengths until full-length genes are gained. *Amplification:* In the last step, recombined full-length genes are amplified by standard PCR.

assemblies by following the instructions of this chapter. The description comprises all steps of the NExT DNA shuffling procedure, including choosing the dUTP to dTTP ratio for the uridine-exchange PCR (Subheadings 3.1 and 3.2), the enzymatic digest and chemical cleavage of the DNA backbone (Subheading 3.3), the fragment purification (Subheading 3.4), and the gene reassembly and amplification (Subheading 3.5), followed by the cloning of the reassembled gene library into a plasmid vector.

2. Materials

2.1. Uridine-Exchange PCR

1. Template DNA.
2. Primers appropriate for amplifying the DNA segments, which need to be recombined.
3. Standard *Taq* DNA polymerase (different suppliers).

4. 10× concentrated PCR buffer (supplied with the used polymerase; see Note 1).
5. Stock solutions (10 or 100 mM) of dATP, dCTP, dGTP, dTTP, and dUTP.
6. Deionized sterile water (see Note 2).
7. 50× Tris-acetate-EDTA buffer (TAE-buffer): 2 M Tris, 1 M acetic acid, 50 mM EDTA. pH 8.0.
8. Ethidium bromide solution: 10 mg/ml ethidium bromide in water (see Note 3).
9. 1% agarose gel: 1% (w/v) agarose in 1× TAE buffer. Melt agarose completely by boiling, e.g., using a microwave. Add ethidium bromide to a final concentration of 0.4 μg/ml (1:25,000) before pouring the gel. Run the gel in 1× TAE-buffer.
10. Gel extraction kit (e.g., GE Healthcare or Qiagen).
11. PCR thermocycler.

2.2. Enzymatic Digest and Chemical Cleavage

1. Uracil-DNA-glycosylase (UDG) from *E. coli* (NEB or Peqlab Biotechnologie GmbH).
2. 5 M NaOH (alternatively piperidine, >99.9% pure from Sigma-Aldrich; see Note 4).

2.3. Fragment Purification

1. QiaexII kit (Qiagen).
2. 3 M Acetic acid, pH 5.2.
3. Deionized sterile water.

2.4. Gene Reassembly and Amplification

1. dNTP mixture of dATP, dCTP, dGTP, and dTTP (10 mM each).
2. Vent DNA polymerase (New England Biolabs).
3. *Taq* DNA polymerase (different suppliers).
4. 10× concentrated PCR buffer (supplied together with Vent DNA polymerase).
5. 25 mM $MgSO_4$.
6. Same primers as used for the Uridine-exchange PCR (see Note 5).
7. Deionized sterile water.
8. 1% agarose gel (see Subheading 2.1).
9. Gel extraction kit (e.g., from GE Healthcare or Qiagen).
10. PCR thermocycler with advanced programming options.

2.5. Cloning

1. Restriction enzymes for excising the desired gene and opening the vector.
2. Appropriate 10× concentrated reaction buffer.

3. Shrimp alkaline phosphatase (Fermentas).
4. 1% agarose gel (see Subheading 2.1).
5. Gel extraction kit (GE Healthcare or Qiagen).
6. PCR cleanup kit (GE Healthcare or Qiagen).
7. T4 DNA ligase (from different suppliers).
8. Appropriate ligation buffer (supplied with enzyme).
9. Competent cells (e.g., *E. coli* XL1-Blue).

3. Methods

3.1. NExTProg

The uridine-exchange PCR (see Subheading 3.2) amplifies the gene pool and incorporates uridine as exchange nucleotide for thymidine. As the sites of exchange-nucleotide incorporation define the fragments and the frequency of incorporation is proportional to the relative amount, the ratio of exchange to standard nucleotide (here, dUTP to dTTP ratio) directly influences fragment sizes. For optimal recombination, a ratio is chosen that minimizes the likelihood that full-length genes without any incorporated exchange nucleotide are synthesized. However, it has to be taken into account that very small fragments (<30 nucleotides) are more difficult to reassemble or might even be partially lost during the fragment purification step (see Note 5).

Instead of optimizing fragmentation experimentally, we propose to use our NExTProg computer program to calculate all possible fragments, their likelihood of occurrence, and their relative distribution dependent on the dUTP to dTTP ratios and the DNA sequence (10, 11) (Fig. 2). The program reads the nucleotide sequence file and dUTP to dTTP values; the complementary strand for a given DNA sequence is automatically generated and taken into account. Upper and lower ranges for the fragment size can be set, and the program calculates the potential loss of material (e.g., because of gel purification) and adjusts the relative likelihood of individual fragments. The calculated results are displayed as a bar chart and can additionally be exported as a tabulated list. For a better overview, the program sums up the probabilities of all fragments of the same length and displays the sums of probabilities versus the length, shown in the program as "%mol". To mirror the resulting picture of gel electrophoresis, the relative mass distribution is also calculated by multiplying the probability of a certain fragment length with its length and is shown in the program as percentage of the total DNA (%mass).

Before using the NExTProg program, there is one important point to consider: the incorporation of uridine versus thymidine by the polymerase in the uridine-exchange PCR. The incorporation-

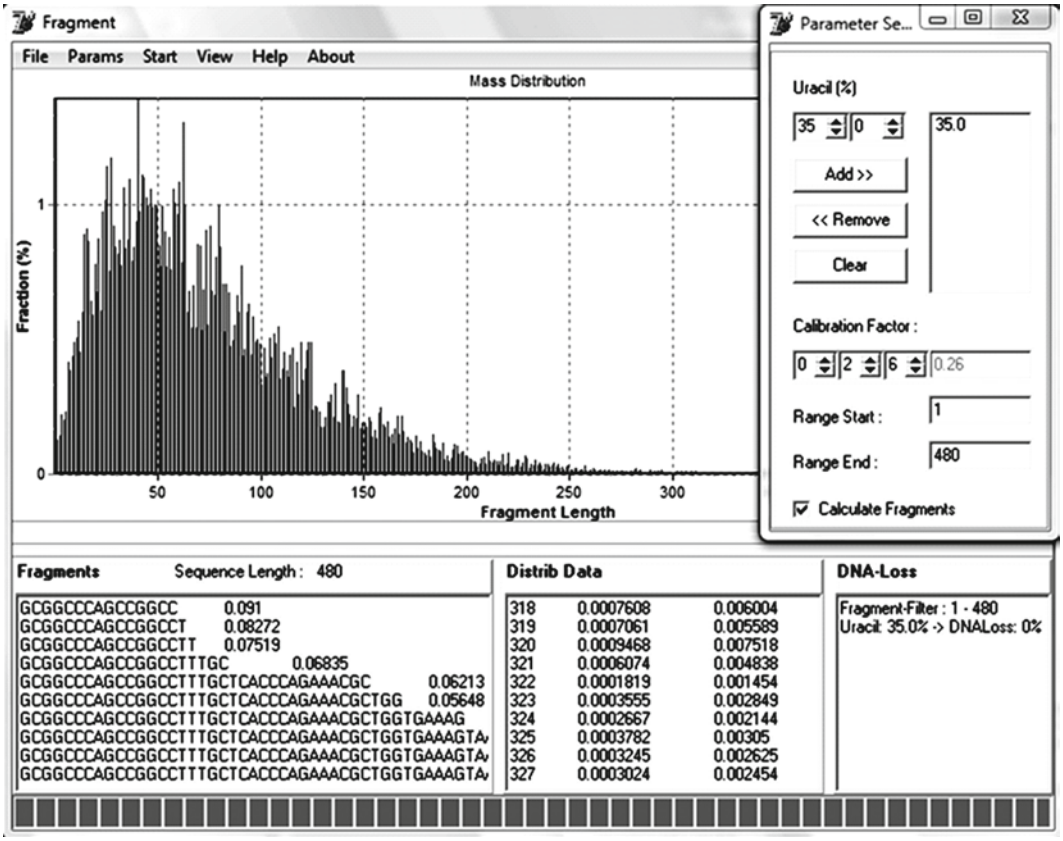


Fig. 2. Graphic user interface of NExTProg 1.0 (10). This program reads a nucleotide sequence file and calculates all possible fragments and their likelihood of generation dependent on the parameter settings (small window). The calculated result is displayed as a bar chart; relative values, length, and sequences of all calculated fragments can be exported as a tabulated list. The view can be changed between the relative mass distribution (shown here) and the relative number of fragments compared to fragments' lengths.

factor depends not only on the dUTP to dTTP ratio but also on the polymerase used, the absolute concentrations of nucleotides, and the buffer (see Note 6). The relative uridine incorporation can be measured experimentally using radioactively labelled dUTP and counting the activity of PCR products, or more indirectly, by quantitatively analyzing the fragmentation pattern as described (11). This value can be set in the program.

To avoid additional experimental work for determining the calibration factor, we recommend following the PCR conditions (standard *Taq* polymerase and 0.4 mM dNTPs) described here (see Note 1), which were used for setting the default value of 0.26 in our program.

3.2. Uridine-Exchange PCR

In the exchange PCR, the gene library is amplified by a standard PCR, in which a fragmentation-defining exchange nucleotide is incorporated together with the four standard nucleotides. dUTP is known to be efficiently incorporated into DNA by various

polymerases (see Note 7) and is therefore advantageous over other possible exchange nucleotides (see Note 8). As mentioned above, the dUTP to dTTP ratio can be adjusted either experimentally or by using the NExTProg program that was developed for this purpose (see Subheading 3.1). For most purposes, ratios of 25–35% dUTP (of the dUTP/dTTP mixture) are appropriate (see Note 6).

1. Dilute 100 mM nucleotide stock solutions with water to 10 mM for dATP, dGTP, and dCTP. Prepare a 10 mM dUTP/dTTP mixture with the chosen dUTP to dTTP ratio (i.e., for 30% dUTP prepare a solution containing 3 mM dUTP and 7 mM dTTP).
2. Prepare a 50-µl PCR reaction mixture containing 20 fmol template DNA (approx. 50 ng for a plasmid of 4,000 bp), 25 pmol of each primer, 5 µl of a 10× concentrated buffer (see Note 1), 5 U *Taq* DNA polymerase, 0.4 mM dATP, dGTP, and dCTP each, and 0.4 mM dUTP/dTTP mixture.
3. Set the thermocycler program as follows: one cycle of 94°C for 2 min; 25 cycles of 30 s denaturation at 92°C, 20 s at an appropriate annealing temperature (see Note 9), and 2 min extension at 72°C (see Note 10); and one final cycle of 72°C for 5 min (see Note 10). Depending on the PCR efficacy, more than one 50-µl reaction might be necessary to obtain enough product (see Note 11).
4. After combining the PCR samples, separate the PCR product from the template by electrophoresis using a 1% agarose gel. Purify the excised gel band using a gel extraction kit and elute with 50 µl water (see Note 12).
5. Optional: determine the DNA concentration of the recovered PCR product by taking the baseline corrected 260 nm value of an absorption spectrum from 220 to 350 nm.
6. Ideally, use all of the DNA for the next steps.

3.3. Enzymatic Digest and Chemical Cleavage

For gene fragmentation, the exchange nucleotide is excised enzymatically, and the DNA backbone of the abasic site is cleaved chemically. In the case of dUTP as exchange nucleotide, the enzyme UDG is used to remove the uracil moiety with high specificity by a nucleophilic attack at the C1' position of the deoxyuridine (12). Chemical cleavage is an alkaline hydrolysis, which is also the underlying principle of Maxam–Gilbert sequencing (13). For an example of a fragmentation, see Fig. 3a.

1. Transfer the purified PCR product (approx. 50 µl) to a PCR tube.
2. For the enzymatic cleavage, add 6 µl of the 10× concentrated UDG buffer and 2 U of *E. coli* UDG to the sample and adjust

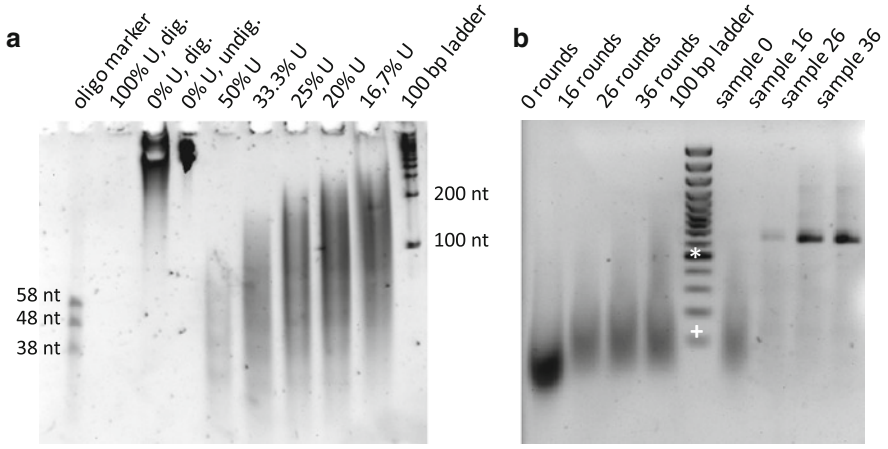


Fig. 3. Overview of the NExT procedure. (**a**) Fragmentation. Ethidium bromide-stained polyacrylamide urea gel showing fragmentation patterns of CAT_Nd10 PCR products obtained with various dUTP/dTTP ratios (%U = c(dUTP)/[c(dUTP) + c(dTTP)] × 100). (**b**) Reassembly and amplification PCR. *Lanes to the left* of the 100 bp DNA ladder show the fragments' increasing size with increasing numbers of denaturation/annealing/elongation cycles (referred to as rounds). *Lanes to the right* show the outcome of the amplification PCR starting from fragments that have been subjected to different numbers of reassembly cycles (*plus sign*, 100 bp; *asterisk*, 500 bp; full-length gene: 627 bp).

the volume to 60 µl with water. Incubate the sample for 1 h at 37°C.

3. To cleave the DNA backbone, add 6.7 µl 5 M NaOH (final concentration 0.5 M) and heat for 30 min at 90°C in a thermocycler with a heated lid (see Note 4).
4. After cleavage with NaOH, do not store the sample over a prolonged period. Proceed with the following step, at least until the pH is neutralized.

3.4. Fragment Purification

The gene fragments can be purified either directly from the reaction solution or over a preparative gel. We prefer direct purification from solution to minimize DNA loss. As we are typically working with fragments of ≤150 bp, we recommend the QiaexII kit (Qiagen; see Note 13).

1. Add the capture buffer included in the kit and neutralize with acetate buffer as recommended in the manufacturer's manual (approx. 20 µl of 3 M acetate buffer).
2. Use 15–20 µl matrix.
3. Wash the matrix twice as described in the manual.
4. For elution, resuspend the matrix in up to 50 µl prewarmed (approx. 50°C) water and incubate for 5 min at 50°C.
5. Filter the sample through a frit to separate the DNA from the matrix. We recommend using a DNA-prep spin column from

which the original silica matrix has been removed. Place the column in a clean 1.5-ml reaction tube and centrifuge for 1 min with maximum speed in a tabletop microcentrifuge (>10,000 rcf) (see Note 14).

6. Optional: repeat steps 4 and 5.

3.5. Gene Reassembly and Amplification

The purified gene fragments are reassembled to full-length genes by an internal primer extension procedure. In this PCR-like reaction with increasing annealing temperatures, the overlapping fragments serve each other as primers and are extended with each cycle until full-length genes are obtained. For high efficacy and low mutational rate, we recommend the use of a proofreading DNA polymerase with long half-life, such as Vent DNA polymerase. Furthermore, *Taq* polymerase adds unwanted nucleotides at the termini (14). The products of this assembly reaction are finally amplified by a standard PCR with end primers. For an example of gene assembly, see Fig. 3b.

1. The amount of purified DNA fragments used for the reassembly can be varied widely. From our experiences, when recombining genes of several hundred base pairs (300–900 bp), approximately 2 µg works well, but less than 1 µg also yields results. If problems with the assembly arise, it is a good idea to increase the amount of fragments. We normally use about one-fourth to one-third of the purified fragments.
2. Prepare a 50-µl reaction mixture containing the DNA fragments, 4 U Vent DNA polymerase (optional: plus 1 U *Taq* DNA Polymerase), 5 µl of the supplied 10× concentrated buffer, 800 µM dATP, dCTP, dGTP, and dTTP each (4 µl of a mixture containing 10 mM of each dNTP), and additional 0.5 mM $MgSO_4$ (see Note 15).
3. The thermocycler program for the reassembly is: one cycle of 94°C for 3 min; 40 cycles of 30 s denaturation at 92°C, annealing over 60 s at increasing temperatures, starting with 30°C and adding 1°C per cycle (preferred cooling ramp 1°C/s), and extension at 72°C for 60 s plus 4 s per cycle; followed by one final cycle of 72°C for 10 min.
4. Use standard PCR conditions to amplify 15 µl of the reassembly product with addition of appropriate primers (25 pmol of each primer, 0.2 mM of each standard dNTP, 25 cycles, 30–60 s elongation time, annealing temperature appropriate for the used primers; see Notes 9, 10 and 16).

3.6. Cloning

Clone the amplified genes via the appropriate restriction sites, using standard methods (15).

4. Notes

1. To minimize the overall error rate, all PCR steps should be performed under conditions optimal for polymerase fidelity. We, therefore, recommend following the manufacturer's instructions for setting the optimal buffer conditions. We used 10× PCR buffer containing 160 mM $(NH_4)_2SO_4$, 670 mM Tris–HCl, pH 8.8 (at 25°C), 15 mM $MgCl_2$, 0.1% (v/v) Tween-20 or 10× PCR buffer containing 100 mM Tris–HCl, pH 9.0 (at 25°C), 500 mM NaCl, 15 mM $MgCl_2$, 1% (v/v) Triton X-100.
2. All solutions should be prepared in deionized and sterile water that has a resistance of 18.2 MΩ.
3. Avoid direct contact with the substance. Ethidium bromide is toxic, cancerogenic, and teratogenic.
4. We found piperidine and NaOH equally well suited for the chemical cleavage. As the toxicity of piperidine exceeds that of NaOH, we recommend using the latter. All work with piperidine should be performed under a fume hood; avoid direct contact. All solutions containing piperidine, e.g., the used capture buffer of the following purification step (see Subheading 3.4), should be handled with care and discharged in a closed tube.
5. In the case of very short DNA segments, it might be advantageous (for easier handling and to avoid extensive loss of very small DNA fragments) to include flanking sequences in the shuffling procedure. After the reassembly PCR, the gene of interest can be amplified with internal primers.
6. The relative amount of exchange nucleotide needed to achieve proper fragmentation depends on length and nucleotide composition of the gene. When using dUTP as exchange nucleotide, difficulties may arise if the genes contain long stretches of thymidine or adenine (thymidine in the complementary strand) or if using a polymerase not sufficiently able to incorporate uridine. In this case, other exchange nucleotides and/or different DNA polymerases can be applied. Efficacy of fragmentation and reassembly can be investigated by monitoring the size of the DNA molecules throughout the procedure on agarose gels. After fragmentation, no full-length gene should be visible. After reassembly, the size of the fragments should have increased and a band of the full-length gene might become detectable.
7. The uridine-exchange PCR can be performed with any polymerase, which is able to sufficiently incorporate the exchange nucleotide. We recommend the use of a standard *Taq* DNA

polymerase. However, proofreading polymerases such as Vent polymerase (New England Biolabs) and nonuracil-stalling mutants of *Pfu* (Stratagene) have been reported to incorporate uridine and might be used instead.

8. An alternative exchange nucleotide/excision enzyme pair is 8-oxo-guanine and FPG (formamidopyridine-DNA glycosylase; New England Biolabs).
9. The appropriate annealing temperature depends on the length and nucleotide composition of the used primers and can be estimated by increasing 4°C for each C or G and 2°C for each A or T.
10. The extension steps should be performed at the polymerase's temperature optimum. The time needed for complete extension of the gene depends on the polymerase, the temperature, and sequence length and can be estimated according to the information of the polymerase instruction leaflet. For the uridine-incorporation PCR, we recommend setting the extension time per cycle to twofold the estimated value. For the very final extension step, use a time exceeding the estimated time by a factor of 5.
11. For a gene length of 500–1,000 bp, approximately 7 µg would be optimal for a trouble-free gene reassembly, but lower amounts work as well.
12. It is necessary to separate the uracil-containing PCR product from the nonuracil-containing template. To avoid dilution of the sample, reload the spin columns up to two times if necessary. We prefer water over Tris buffer for the elution to have more defined buffer conditions in the following reaction steps. The use of warm water as well as an incubation time of more than 1 min may improve DNA recovery.
13. The QiaexII kit is recommended by the manufacturer only for fragments larger than 40 bp; however, we were able to recover fragments down to approximately 20 bases in low amounts, possibly those small ssDNA fragments hybridized with larger ssDNA fragments before purification. Other PCR cleanup kits can be used depending on the desired fragment sizes and the technical specifications of the kit. To ensure that the DNA binds to the matrix, the pH of the sample has to be adjusted according to the manufacturer's instructions.
14. The use of a spin column guarantees that the matrix, which would interfere with following reaction steps, is completely removed from the sample.
15. Besides increasing the yield of PCR product, the addition of a standard *Taq* DNA polymerase might slightly increase the overall error rate. For optimal yield, titration of the $MgSO_4$ concentration (up to 8 mM) might be necessary.

16. If the efficacy of the amplification PCR is not satisfactory, the reassembly might not have been completed. In this case, the amplification PCR should be started with ten additional cycles with reduced annealing temperature.

Acknowledgments

We thank Susanne Knall for technical assistance, Gregor Zipf for coding NExTProg, and Hubert Bernauer for fruitful discussions.

References

1. Gaida, A., Hagemann, U. B., Mattay, D., Räuber, C., Müller, K. M., and Arndt, K. M. (2009) Selection of peptides interfering with protein-protein interaction. *Methods in Molecular Biology* **535**, 263–91.
2. Jouaux, E. M., Schmidtkunz, K., Müller, K. M., and Arndt, K. M. (2008) Targeting the c-Myc coiled coil with interfering peptides. *Journal of Peptide Science* **14**, 1022–31.
3. Mason, J. M., Müller, K. M., and Arndt, K. M. (2007) Positive aspects of negative design: simultaneous selection of specificity and interaction stability. *Biochemistry* **46**, 4804–14.
4. Amstutz, P., Forrer, P., Zahnd, C., and Plückthun, A. (2001) In vitro display technologies: novel developments and applications. *Current Opinion in Biotechnology* **12**, 400–5.
5. Wong, T. S., Roccatano, D., Zacharias, M., and Schwaneberg, U. (2006) A statistical analysis of random mutagenesis methods used for directed protein evolution. *Journal of Molecular Biology* **355**, 858–71.
6. Stemmer, W. P. (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 10747–51.
7. Stemmer, W. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–91.
8. Zhao, H., Giver, L., Shao, Z., Affholter, J. A., and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature Biotechnology* **16**, 258–61.
9. Shao, Z., Zhao, H., Giver, L., and Arnold, F. H. (1998) Random-priming in vitro recombination: an effective tool for directed evolution. *Nucleic Acids Research* **26**, 681–3.
10. Müller, K. M. and Zipf, G. (2005) NExTProg: http://www.molbiotech.uni-freiburg.de/next.
11. Müller, K. M., Stebel, S. C., Knall, S., Zipf, G., Bernauer, H. S., and Arndt, K. M. (2005) Nucleotide exchange and excision technology (NExT) DNA shuffling: a robust method for DNA fragmentation and directed evolution. *Nucleic Acids Research* **33**, e117.
12. Drohat, A. C., Jagadeesh, J., Ferguson, E., and Stivers, J. T. (1999) Role of electrophilic and general base catalysis in the mechanism of Escherichia coli uracil DNA glycosylase. *Biochemistry* **38**, 11866–75.
13. Maxam, A. M. and Gilbert, W. (1977) A new method for DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 560–4.
14. Clark, J. M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Research* **16**, 9677–86.
15. Sambrook, J. and Russell, D. W. (2001) Molecular cloning: a laboratory manual. 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

INDEX

A

Adapters 44, 97, 136–146, 261–262
Aerosols 5, 7, 9, 13, 39, 116
Affinity maturation 283, 295–296
Agarose gel electrophoresis 48–51, 108, 145, 163, 199, 224, 249, 280, 326
Alkaline phosphatase 247–249, 255, 291, 299, 336–337
Allele-specific PCR 166, 172–173
Alu elements 221–222, 226–227
Alu PCR 221–228
Antibiotic susceptibility testing 248, 251–252
aPCR. *See* Asynchronous PCR
Archaeal DNA polymerases 18
Array hybridization 240
aSP-PCR. *See* Asymmetric solid phase PCR
Asymmetric overlap extension (AOE) PCR 277–279, 282
Asymmetric PCR 237–239, 242, 277–278, 281, 282
Asymmetric solid phase PCR (aSP-PCR) 257–258
Asynchronous PCR (aPCR) 231–242

B

Bacterial ligase 248, 252–256
BeadArray 114
Blocking oligonucleotides 265–272
Blocking primers 271, 272
Blocking probes 270–271
Bridge PCR 258, 259

C

Cap analysis of gene expression (CAGE) 147–163
Capillary electrophoresis 182, 193–194, 223–225, 269–270
CapSelect 43
CapTrapper 54
Cap-Trapping 154–155
Cassette ligation-anchored PCR 26
cDNA amplification 91–109
cDNA synthesis 92–94, 97–100, 116, 126, 137, 139, 140, 147–148, 196, 241, 303
Cell-free protein synthesis 309–310, 313, 315–316
Chemical cleavage 335, 336, 339–340, 342
Chimeric genes 277
Chimeric oligomer 232
ChIP. *See* Chromatin immunoprecipitation
ChIP-chip 307
Chromatin 177–190, 195, 203
Chromatin immunoprecipitation (ChIP) 307
Chromosomal breakpoints 30
Circularization 26
Clamp region 58, 64, 65, 67, 69
Cloning 17, 26–29, 32, 37–40, 76, 79–84, 135, 139, 165, 194, 195, 245, 270, 279, 281, 288, 298, 335–337, 341
CODEHOP. *See* Consensus–Degenerate Hybrid Oligonucleotide Primer
Competitive probes 265
Concatemerisation 269
Consensus clamp 58–59, 63–67, 69
Consensus–Degenerate Hybrid Oligonucleotide Primer (CODEHOP) 57–71
Consensus primer 58
Conserved primers 265, 270
Contig construction 17
Copy number variation 193–194
CpG methylation 195, 196, 213
Ct values 104, 107–108, 128–131, 173, 174, 208, 214–215, 236, 237, 270

D

Deadenylated ligase 250, 252
Deamination 18
Decontamination 9, 11–13
Deep CAGE 147–163
Degenerate primer 58
Denaturing gradient gel electrophoresis (DGGE) 270
Denaturing high-performance liquid chromatography (DHPLC) 269–270
Deoxyribonuclease I (DNaseI) 178, 180, 187, 195, 197, 198, 202–204, 287
DGGE. *See* Denaturing gradient gel electrophoresis
Digital PCR 166
Dinucleotide repeats 187, 268–269

Daniel J. Park (ed.), *PCR Protocols*, Methods in Molecular Biology, vol. 687,
DOI 10.1007/978-1-60761-944-4, © Springer Science+Business Media, LLC 2011

Directed evolution 165, 286, 333–344
DNA fingerprinting 48, 54, 221
DNA library 26, 284, 285, 291–292, 309, 312, 321
DNA polymerase 17–23, 27–28, 31, 39–40, 45–52, 54, 78, 79, 83–85, 95, 100, 126, 130, 137, 138, 140, 142, 150, 166, 168, 171, 172, 178, 188, 189, 195, 228, 231, 259, 262, 263, 267, 278, 300–301, 312, 341–343
DNA–protein interaction 307–316
DNaseI. *See* Deoxyribonuclease I
DNA substrates 245–256

E

EcoP15I 136, 138, 139, 141, 144–145, 147–150, 158–159, 163
ELISA. *See* Enzyme-linked immunosorbent assay
Emulsion PCR 165–174, 257, 259, 261–262, 308, 312, 316
End repair priming 26, 28–29
Enhanced solid phase (ESP) PCR 258–263
Enzyme-linked immunosorbent assay (ELISA) 246, 255, 298, 320, 322, 323, 327–329
Error-prone PCR 166, 285, 286, 297–298, 302, 333–334
Evolving proteins 283–304
Exchange nucleotides 333–344
Exonuclease 44, 108, 178–179, 195, 267, 272, 281, 291
3' to 5' Exonuclease 267, 281
Exonuclease I 93–98, 106, 108, 109
ExoSAP 215, 310, 314
Expressed sequence tag (EST) analysis 107, 135–136
Expression profiling 91, 114, 130, 136, 147, 329
Extension MLPA 195, 202

F

Flanking sequence 26, 28, 43–55, 58, 63–65, 342
Flow cytometry 259–263, 308, 310, 313, 314
Fluor 262–263
Fluorescent 106, 116–117, 124, 125, 145, 180, 182, 184–187, 193, 197, 208, 209, 213, 224, 259, 262, 263, 269, 270, 310, 313, 321
Fusion DNA polymerase 17–23, 300

G

Gel electrophoresis 20, 21, 44, 48–51, 61, 100, 101, 107, 108, 142, 145, 163, 199, 222, 224, 237, 239, 241, 249, 269, 270, 280, 310, 312, 315, 326, 337, 339
Gel-mobility shift assay 308, 311, 314–316
Gene expression 91, 92, 95, 102–104, 106–108, 113, 114, 127, 128, 130, 131, 135–136, 145, 147–163, 187, 232, 235, 307, 320–323, 327–330
General primer 44, 59, 268
Gene-specific primer 44–47, 52, 100, 184, 284, 291
Genetic library on beads (GLOBE) 307–317
Genome mapping 17, 162
Genome-wide analysis 26, 91, 95
Genomic DNA extraction 19, 22, 30–33, 166, 223, 224
Genomic DNA fragments 19, 30, 222
Genomic insertion sites 25–41
GLOBE. *See* Genetic library on beads

H

Hairpin 20, 26, 27, 52, 269
Haplotype analysis 165–174
Haplotyping 165–174
Heavy methyl assay 208
High performance liquid chromatography (HPLC) 20, 22, 93, 137, 182, 269–270, 272
High resolution melting (HRM) 207–216
curves 270
High throughput SuperSAGE (HT SuperSAGE) 135–146
Homologous genes 59, 335
Hot start 48–51, 54–55, 68, 93, 99, 101–103, 138, 142, 168, 172, 209, 213, 247, 249, 259, 262, 292, 297
HPLC. *See* High performance liquid chromatography
HRM. *See* High resolution melting
HT SuperSAGE. *See* High throughput SuperSAGE
Hybrid primer 45, 49, 52, 54, 57–59, 214, 231
Hybrid selection 166

I

Immuno PCR 246
Insertion mutagenesis 26, 48, 54
Intron-exon junctions 30
Invader assay 114
Inverse PCR 26, 27, 43
Inverted end 267
In vitro ligand evolution 307
In vitro protein synthesis 316
In vitro splicing 277, 279
In vitro transcription (IVT) 92, 103, 285, 287, 288, 290–292, 296, 298, 303, 308, 313–315
In vitro translation 286, 289–295, 300, 308, 310, 313, 314
In vivo footprinting 187–188
IVT. *See In vitro* transcription

K

Klenow fragment 309, 311, 312

L

Laboratory setup 3–14
Lariat-dependent Nested (LaNe) PCR 44, 53

Lariat-dependent Nested (LaNe) RAGE 44, 45, 47, 48, 51, 52, 54
Lariat structure 44, 45, 49, 52, 54
Libraries 26, 136, 139, 259, 261, 283–305, 309, 312, 321, 325, 333–335, 338
Ligase 31, 34, 36, 43, 44, 137, 138, 140–142, 150, 158, 181, 184, 195–197, 199, 201–202, 247, 248, 250–256, 288, 297, 337
Ligase chain reaction (LCR) 31
Ligase-mediated PCR 26, 28–29, 43
Limiting dilution 166
Linear PNA beacon probes 232, 234
Locked nucleic acid (LNA) 124, 267
Long range PCR 17–23
Long SAGE 135
Long terminal repeat 27

M

Mapping 26, 39, 40, 152, 161–162
Massively parallel sequencing signatures (MPSS) 135
Maxam-Gilbert sequencing 339
Megaprimer PCR 277
Melting temperature 20, 52, 66, 67, 208, 241, 267, 329
Methylation 177, 195, 196, 198, 201, 207–216
Methylation sensitive high resolution melt analysis (MS-HRM) 208, 209, 210, 213, 215
MethyLight Real-time PCR assay 207
Microarray 91–109, 114, 136, 147, 200, 235, 237, 240, 307, 320–321, 323–326, 329
Micrococcal nuclease (Mnase) 177–180
MicroRNA 113–132
MLPA. *See* Multiplex ligation-dependent probe amplification
Mnase. *See* Micrococcal nuclease
MPSS. *See* Massively parallel sequencing signatures
mRNA 46–48, 53, 54, 95–96, 104, 107, 109, 114, 128, 196–197, 207, 283–285, 287, 290–292, 295, 296, 299, 301, 321, 327, 328
MS-HRM. *See* Methylation sensitive high resolution melt analysis
Multiple site mutagenesis 277
Multiplex ligation-dependent probe amplification (MLPA) 193–204
Multiplex PCR 246, 259
MuLV. *See* Murine leukemia virus
Murine leukemia virus ((MuLV) 26, 27
Mutagenesis 25, 26, 48, 54, 84, 85, 277–282, 334
Mutagenic nucleotide analogue 76, 79

N

NAD-dependent DNA ligase 253, 254
Nested PCR 43–55
New generation sequencing (NGS) 78, 135, 136, 270
NGS. *See* New generation sequencing
Non-coding RNA 113
Non-radioactive labelling 182
Non-target sequences 265, 269, 271
Nucleotide exchange and excision technology (NExT) DNA shuffling 333–344

O

One-sided PCR 44
Overlap extension (OE) PCR 277–282

P

Padlock probe 114
PAGE. *See* Polyacrylamide gel electrophoresis
PAP-LMPCR. *See* Pyrophosphorolysis activated polymerization ligation-mediated PCR
Paralogs 59, 60, 71
PCR carryover 11, 171
PCR contamination 3–13, 39, 166, 171, 207, 255, 262
PCR walking 43–44
Peptide nucleic acid (PNA) 232–238, 240, 267, 269
Phi29 DNA polymerase 83
PNA. *See* Peptide nucleic acid
Polony haplotyping 166
Polyacrylamide gel electrophoresis (PAGE) ..... .125, 142, 143, 145, 150, 160, 161, 222, 237, 241, 242, 272, 312
Poly (dA) tailing 95
Primer design software 52, 60, 61, 65–67, 109, 214, 235, 236, 241, 268
Probe design 195, 197, 201, 320, 324–325, 328, 329
Proofreading DNA polymerase 17–18, 281, 341–343
Protein-DNA binding specificities 320
Pyrophosphorolysis activated polymerization ligation-mediated PCR (PAP-LMPCR) 177–191

Q

qPCR. *See* Quantitative PCR
qRT-PCR. *See* Quantitative reverse transcription PCR
Quantitative PCR (qPCR) 10, 116–117, 123, 125–130, 193, 231–233, 236–237, 241, 270
Quantitative reverse transcription PCR (qRT-PCR) 147

R

RACE. *See* Rapid amplification of cDNA ends
RACE LaNe 44, 46, 48–50, 52–54
RAGE. *See* Rapid amplification of genomic DNA ends
Random priming recombination (RPR)..... 334
Rapid amplification of cDNA ends (RACE).....26, 43–44, 46, 48–50, 52–54, 136
Rapid amplification of genomic DNA ends (RAGE).....26, 43–45, 47, 48, 51, 52, 54
Rare sequences..... 265–266
RCA. *See* Rolling circle amplification
Real-time PCR (RT-PCR) 6, 92, 94–95, 100, 102–108, 114–118, 121, 123–125, 127, 130, 173, 207–208, 232, 234, 247, 252, 254–255, 271
Regulatory region 30, 197, 321, 329
Repeat sequences 78
Reporter probes 270
Restriction endonuclease 44, 195–196, 285, 288, 291
Ribosome display..... 283, 285–286, 288, 291, 304–305, 333
RNA preparation.....115, 117–120
RNaseH..... 44, 46, 48–50, 93, 99, 149
RNaseI 147–149, 154
Rnase inhibitors.....93, 97, 104–105, 109, 116, 121, 122
RNA-Seq 135
Rolling circle amplification (RCA) 78, 82–83
RPR. *See* Random priming recombination
RT-PCR. *See* Real-time PCR

S

SAGE. *See* Serial analysis of gene expression
SAM. *See* Sequencing aided by mutagenesis
Secondary structure 20, 22, 52, 279, 329
SELEX-SAGE 307
Sequencing aided by mutagenesis (SAM) 75–76, 78–79, 81, 84
Serial analysis of gene expression (SAGE) 135–136, 139, 144, 307
S30 extract..... 284–285, 289, 290, 293, 299–300, 310, 313
Shuffling.....286, 334–235, 342
Single cell isolation.....92–93, 95–97
Single cell microarray 95
Single nucleotide polymorphism 26
Single primer amplification (SPA) 320, 326
Single specific primer PCR 26
Single strand conformational polymorphism (SSCP) 270
Site-directed mutagenesis..... 277–279
Site-saturated mutagenesis 334
Small RNA.....113, 117, 119, 125
Smeared products 44
Sodium bisulfite modification 209, 212, 214
Solid phase PCR 257–258
Solid support 257–263
Southern blotting 37–38, 40, 193
SPA. *See* Single primer amplification
Spacer..... 185, 190, 267, 270, 272, 285, 291, 297
Spike RNAs..... 94, 95, 97–98, 103–104, 106–109
SplinkBlunt-PCR.....27–32, 35–38, 40
Splinkerette PCR 26–27, 43
SplinkTA-PCR 26–27, 29
Staggered extension process (StEP) 334
Stem-loop..... 114, 130
StEP. *See* Staggered extension process
Stoffel fragment..... 195, 197–198, 202, 267
Stringency 53, 269, 302
SuperSAGE 135–136, 139, 144, 145
Systematic evolution of ligands by exponential enrichment (SELEX) 307–308

T

TA cloning 32, 37, 40, 281
TAIL PCR. *See* Thermal asymmetric interlaced PCR
Taqman 114–118, 120–124, 128–131, 208, 232–236, 240–241, 270
Target sequences..... 8, 19, 203, 265, 269, 314
Terminal transferase 46–47, 49–50, 52, 53
Tether sequence 263, 285, 291
Thermal asymmetric interlaced PCR (TAIL PCR)..... 43
T-linker PCR 26, 39
Transcriptional start site (TSS) 147, 187
Transcription factor 114, 117, 307–310, 313–316, 320, 323, 324, 327–329
Transformation..... 38–41, 85, 283–284, 298, 304
Transposon 25, 26, 30, 48, 54
T7 RNA polymerase 95, 288, 291–292, 310
TSS. *See* Transcriptional start site

U

UDG. *See* Uracil DNA glycosylase
Universal Fast Walking..... 43–44, 54
Untranslated region 114, 169
Uracil DNA glycosylase (UDG).....336, 339–340
Uracil poisoning 18
Uridine-exchange PCR 335–339, 342

V

Vectorette PCR 26
VISA technique..... 26